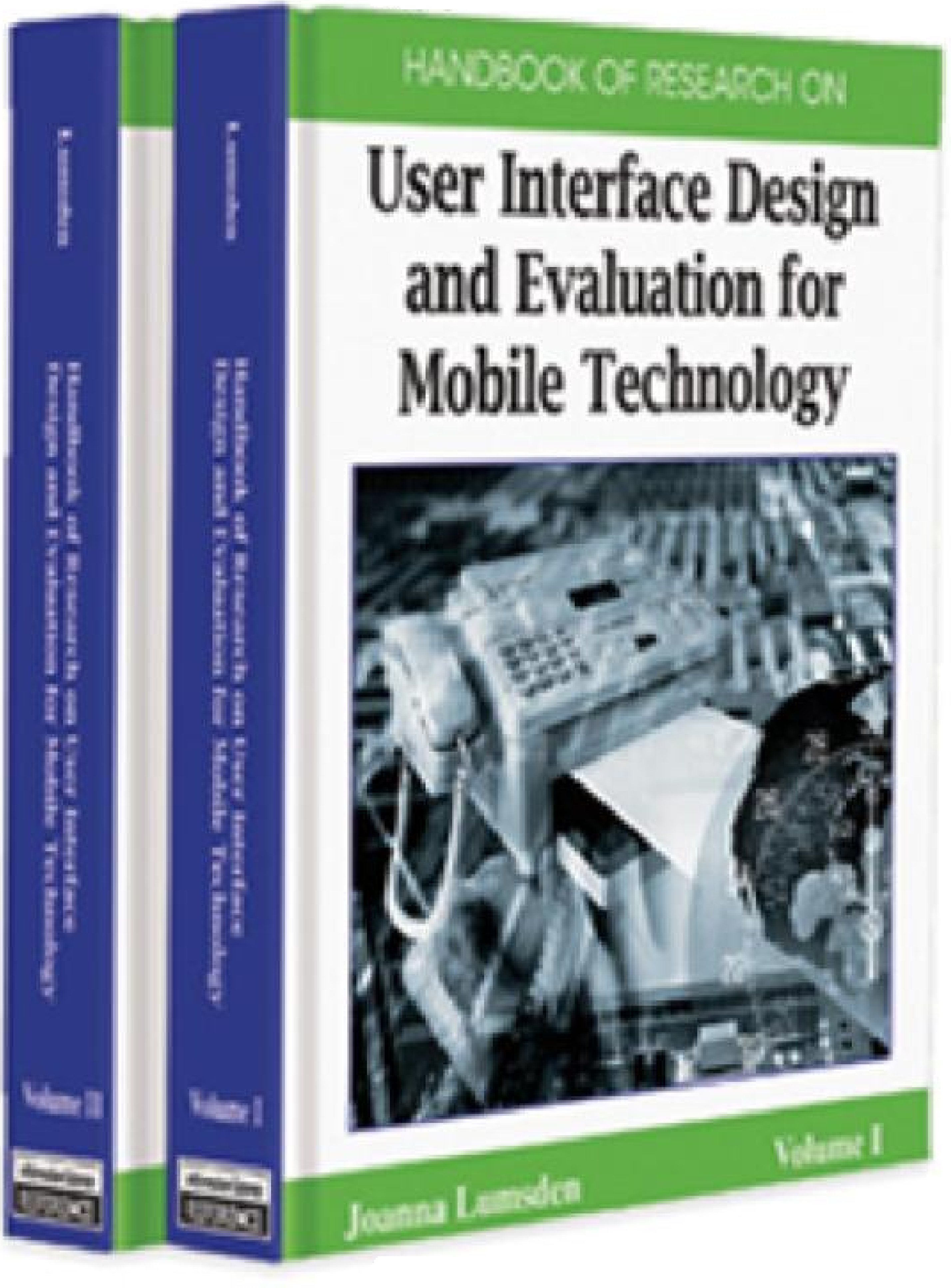

HANDBOOK OF RESEARCH ON
User Interface Design and Evaluation for Mobile Technology
Volume I
Joanna Lumsden
Volume II
Volume I

Handbook of Research on User Interface Design and Evaluation for Mobile Technology

Volume I

Joanna Lumsden
National Research Council of Canada
Institute for Information Technology – e-Business, Canada

INFORMATION SCIENCE REFERENCE
Hershey • New York

Acquisitions Editor: Kristin Klinger
Development Editor: Kristin Roth
Senior Managing Editor: Jennifer Neidig
Managing Editor: Sara Reed
Copy Editor: Joy Langel, Katie Smalley, and Angela Thor
Typesetter: Jeff Ash
Cover Design: Lisa Tosheff
Printed at: Yurchak Printing Inc.

Published in the United States of America by
Information Science Reference (an imprint of IGI Global)
701 E. Chocolate Avenue, Suite 200
Hershey PA 17033
Tel: 717-533-8845
Fax: 717-533-8661
E-mail: cust@igi-global.com
Web site: http://www.igi-global.com

and in the United Kingdom by
Information Science Reference (an imprint of IGI Global)
3 Henrietta Street
Covent Garden
London WC2E 8LU
Tel: 44 20 7240 0856
Fax: 44 20 7379 0609
Web site: http://www.eurospanonline.com

Library of Congress Cataloging-in-Publication Data

Handbook of research on user interface design and evaluation for mobile technology / Joanna Lumsden, editor.

p. cm.

Summary: "This book provides students, researchers, educators, and practitioners with a compendium of research on the key issues surrounding the design and evaluation of mobile user interfaces, such as the physical environment and social context in which a device is being used and the impact of multitasking behavior typically exhibited by mobile-device users"--Provided by publisher.

Includes bibliographical references and index.

ISBN 978-1-59904-871-0 (hardcover) -- ISBN 978-1-59904-872-7 (ebook)

1. Mobile computing--Handbooks, manuals, etc. 2. Human-computer interaction--Handbooks, manuals, etc. 3. User interfaces (Computer systems)--Handbooks, manuals, etc. I. Lumsden, Joanna.

QA76.59.H36 2008

004.165--dc22

2007024493

British Cataloguing in Publication Data
A Cataloguing in Publication record for this book is available from the British Library.

Editorial Advisory Board

Table of Contents

Section IV
Evaluation Techniques for Mobile Technologies

Section V
Case Studies

Detailed Table of Contents

The results of this study suggest that one-handed use should be a serious consideration when designing mobile phones. Prompted by their initial study, the authors conducted a more extensive survey of mobile device use to highlight one-handed usage patterns relative to task and device type. The results of this study show both current usage patterns and preferred usages patterns, and from this, the authors conclude that designers should make one-handed usability a priority in order to bridge the gap between current and desired usage patterns. Finally, the chapter reports on the results of a thumb movement study designed to investigate the extent to which user performance was influenced by device size, task region, and movement direction. On the basis of this final study, the authors present guidelines concerning the use of thumb movement in mobile interaction design.

Usability alone does not determine user acceptance of mobile services. This chapter introduces the technology acceptance model for mobile services (TAMM), which models the way in which a complex set of user acceptance factors affect user acceptance of mobile services. Based on field trials of several mobile services with more than 200 test users, the TAMM is designed to be used as a design and evaluation framework when creating new mobile services. Supporting the identification of issues that should be focused on during the design of new mobile services to ensure user acceptance, the motivation of the TAMM is different than that of the original TAM, which was built to *explain* user acceptance. The TAMM defines four main user acceptance factor (perceived value, perceived ease of use, trust, and perceived ease of adoption) and this chapter discusses design implications associated with each.

This chapter reflects on an examination of the complex relationship between user characteristics or diversity and menu navigation performance on mobile devices. The reported study considered the effect of age, gender, cognitive factors, motivational factors, and prior technological experience in terms of people's ability to effectively navigate menu structures on mobile phones. On the basis of a detailed analysis of individual interaction patterns to determine the effectiveness and efficiency of individuals' menu navigation, the authors discuss implications for the design of mobile phones. Most noticeably, the results show that although the use of small screen devices poses certain difficulties for *all* users, children and middle-aged adults were seen to be very sensitive to the cognitive demands imposed by current mobile phone designs.

Usability studies often recruit participants from university student populations. This chapter reports on the benefits or results to be gained from including users with cognitive difficulties, in this case, learning-disabled children, a disregarded user group with respect to mobile technology design. By comparing learning-disabled children's interaction with mobile phones to that of "average" children and university students, the authors are able to demonstrate the qualitatively and quantitatively different insights into the impact of specific design decisions brought about by including this disregarded user group. The authors suggest that, when participants are restricted to university students, mobile technology evaluations typically fail to observe the full extent to which characteristics of the user interface impact on ease of use.

This chapter introduces the adaptive mobile learning management system (AM-LMS) platform, which represents an adaptive environment that continually matches the needs and requirements of individual learners in a mobile learning environment to a mobile device's use of remote learning resources. The chapter discusses the concept of, and research trends associated with, mobile learnin,g as well as classifications of learning styles and how these relate to the AM-LMS platform, before describing the structure of the AM-LMS itself. The AM-LMS platform presents an ability to take advantage of the unique features of mobile devices combined with the ability to support individualized learning and learner-centered education.

Designing graphical user interfaces for effective use on mobile devices is challenging given the extent to which device capabilities differ and contexts, combined with user preferences, continually change. The authors of this chapter propose that the solution to this problem lies in using a single, abstract, user interface description that is then used to automatically generate user interfaces for different devices. Although there are a number of techniques possible for translating an abstract specification to a concrete interface, this chapter suggests that, by using indirect generation, it is possible to perform run-time analysis of computer-human interaction and application of artificial intelligence techniques, and thereby increase the resulting graphical user interface's performance and usability. The authors present their proposal for an indirect generation approach.

Meeting the demand for increasingly sophisticated applications and services poses new challenges for designers of mobile applications. The authors of this chapter suggest that careful adoption of intelligent user interfaces may offer a practical approach for evolving mobile services. To this end, using the intelligent agent paradigm for illustration purposes, this chapter discusses in detail the issues associated with adopting intelligent techniques in mobile applications.

This chapter introduces the potential for informal prototyping tools to speed up the early stage design of mobile interactions. The authors present two tools that address the early stage design of speech-based and location-enhanced interactions as proofs of concept for informal prototyping tools for mobile interactions. The chapter highlights the use of storyboarding and Wizard of Oz (WOz) testing, and discusses how these can be applied. The authors report on a case study to illustrate the iterative design of a location-aware application.

hardware keyboard designs, different on-screen keyboard layouts, handwriting-based approaches, and more novel approaches such as gestures) and reflects on the nature of the evaluations that have been conducted to assess their validity. The authors discuss criteria for acceptance of new text entry techniques, and comment on how market perceptions can overrule laboratory successes. The chapter concludes with some interesting yet intentionally controversial statements to encourage the reader to consider the future of text entry on mobile devices.

Entering English text into a mobile phone can be challenging, but entering Chinese characters (of which there are thousands) using a mobile phone keypad is much more difficult. This chapter presents work, detailed on a step-by-step basis, that was undertaken to redesign the keypad graphics (that is, the symbols printed on the keys as the legends for Chinese strokes) for the Motorola iTap™ stroke-based input solution. Presented as a case study, this chapter introduces and compares the original iTap™ solution and the Pinyin method. It describes an alternative design, and details two longitudinal studies of the proposed design conducted in the USA and China. The chapter also discusses the simplification of the proposed design to allow it to fit on smaller keypads, and outlines its subsequent evaluation.

The authors of this chapter posit that the use of speech-based interaction, when combined with other modes of interaction, can enhance user experience with mobile technology. This chapter describes a prototype system that supports client-side, voice-enabled applications on mobile devices. The authors reflect on the design issues they faced, and evaluation methods they employed, during the development of a voice-enabled user interface for a mobile device. Interestingly, they discuss the need to evaluate the user interface and speech-recognizer independently to avoid mixing of data leading to inconclusive results, and describe how they achieved this separation when evaluating their system.

Multimodal user interfaces support interaction with a system via multiple human-computer communication channels (or modalities) and, as such, are particularly beneficial for mobile devices in terms of their ability to circumvent the complexities introduced by the limited form-factor of mobile devices and the varied contexts in which mobile devices are used. Taking a speech-centric view, this chapter surveys multimodal user interface design for mobile technology. Based on a selection of carefully chosen case studies, the authors discuss the main issues related to speech-centric multimodal user interfaces for mobile devices.

suggest that individuals with disabilities and domain experts be involved throughout the development process. Although this presents its own set of challenges, many strategies have successfully been used to overcome the difficulties and maximize the contributions of users and experts alike. Guidelines based on these strategies are discussed and are illustrated with real examples from active research projects.

Faced with an aging population, there is a challenging need to design technologies to help older adults remain independent and preserve their quality of life. Current research is attempting to address this challenge by means of assistive technologies based on mobile devices such as personal digital assistants and cell phones, but the question remains as to whether or not older people can use such technologies effectively because of age-related problems. This chapter discusses issues surrounding the design, implementation, and evaluation of mobile applications for older adults, highlighting the unique challenges posed by, and best practices for working with and designing for, this population.

This chapter discusses issues that arise when designing and evaluating mobile software for sensitive situations, where access to end-users is extremely restricted and traditional design and evaluation methods are inappropriate and/or not viable. Specifically, this chapter focuses on the use of technology to support adolescents in mental health care settings, and highlights the corresponding contraints (both practical and ethical) that affect approaches to design and evaluation. The authors present some design recommendations for technological interventions of this nature, and suggest methods to maximise the value of evaluations conducted under such restricted situations. The issues outlined in this chapter are illustrated, or made tangible, by means of a case study relating the design and evaluation of a mobile phone-based "mood diary" application for use in clinical situations by adolescents undergoing mental health interventions.

Radio frequency identification (RFID) has the potential to enhance assistive technologies for mobile users; for example, RFID-based applications may support users with visual impairment by providing information on their current location and physical surroundings. This chapter discusses an extension to a mobile application development tool to include support for designing RFID-based applications. The authors describe an RFID-enabled location-aware tour-guide they developed using the development environment, highlighting the main concepts of the interaction modalities they designed to support visually impaired users. The authors describe a field evaluation of this application, and discuss the results which illustrate the costs, limits, strengths, and benefits of the new technology.

This chapter looks at the concept of using non-speech audio for building interfaces to support visually impaired users. Specifically, it focuses on "unconventional" mechanisms by which to use non-speech audio to enable users with visual impairments to conceptualize spatial information. The authors report on studies they conducted towards meeting this goal, from which they were able to observe that sound effects, such as reverberation and reflection, enhance users' ability to localize pattern-associated sounds, and that "natural" sounds (as opposed to "artificial' sounds) better supported users' conceptualization of spatial information.

This chapter reports on the process undertaken to design an interface for a mobile navigational aid for blind pedestrians. It presents a set of rules for producing route descriptions for blind users, and outlines the method (based on user- and activity-centred approaches) by which the rules were established. The authors reflect on the state of the art of wearable navigational aids, and present their approach to providing improved functional specifications for designing rules for producing verbal instructions and information for blind pedestrians.

This chapter reflects on design trends for interfaces to smart wheelchairs, and highlights the need to take into account the similarity between smart wheelchairs and autonomous mobile robots as well as the specific restrictions imposed by the users and task. The authors discuss the main aspects of the user-wheelchair interface, including the need for an adaptive design approach, and a case study is used to illustrate the design of user, context, and task models to support an intelligent adaptable interface. The chapter also includes discussion of the influence of new navigation models in the design of the user interface.

This chapter discusses, from a theoretical perspective, the concepts and issues involved in evaluating mobile human-computer interaction. The chapter assumes that "usability" is not a feature of a product, but rather it is the consequence of a given user employing a given product to perform a given activity in a given environment. The author argues that mobile context-of-use must be considered in order to assess usability as a concept of "fitness-for-purpose", and proposes a usability evaluation process specifically focused on mobile technology evaluation.

The variability of users, uses, and contexts makes evaluating mobile applications challenging. This chapter outlines various traditional usability evaluation methods and discusses methodological variations to these approaches which make them better suited for evaluating usability aspects of mobile devices and applications. The authors suggest that a combination of both field evaluation methods and traditional laboratory testing should be used to address different phases in the user-centered design and development process for mobile technologies and introduce their "real world lab" concept as means by which to achieve both.

Mobile user interfaces are often used in dynamic environments, with the result that user experiences can vary substantially. Consequently, this increases the complexity associated with effective evaluation of such interfaces. In response to this challenge, this chapter presents a framework for the systematic selection, combination, and tailoring of evaluation methods for context-aware applications based on seven evaluation constraints (the stage of development, the design complexity, and the purpose, participants, setting, duration, and cost of evaluation). This chapter describes how the framework was applied to the evaluation of a context-aware mobile interface for use by the police and how, as a result of this case study, the authors were able to derive specific guidelines for selecting evaluation methods and were able to reflect on the relationship between the mobile context and the user experience.

For desktop-based applications, heuristic evaluation is recognized as a cost-effective mechanism by which to identify a large proportion of usability flaws with limited resource investment. On the other hand, limited screen real estate, divided user attention, and elaborate contextual factors pose complex problems for usability assessment of mobile applications. This chapter describes a modified collection of usability heuristics, systematically derived from extensive literature and empirically validated, that are designed to be appropriate for evaluation of mobile technologies.

This chapter introduces and describes the concept of Wizard of Oz studies. It discusses the use of such studies to evaluate mobile technologies, illustrating the issues raised with a case study. The author presents both a taxonomy for Wizard of Oz studies and a set of guidelines regarding the considerations that are essential when planning Wizard of Oz studies for mobile applications. The author argues that well planned Wizard of Oz studies can provide valuable information about user behaviour and experience that might otherwise be difficult to establish. Furthermore, this chapter suggests that the extent of use of Wizard of Oz studies will likely increase as the complexity of mobile systems increases and establishes greater demands for low-cost methods for early investigation and evaluation.

This chapter proposes a framework and model for identifying, organizing, and classifying usability factors of mobile phones. The conceptual framework incorporates multiple views (including user, product, interaction, dynamic, and execution views) to explain different aspects of the interaction between users and mobile phones, and then describes usability factors in terms of these views. The authors describe a hierarchical model for the classification of usability factors in terms of goal-means relationships, which they developed based on the framework. The chapter outlines two case studies used to verify the usefulness of the framework and model, and presents a set of checklists designed to enhance their practicality for use.

This chapter raises the question of whether or not lab-based usability tests of mobile technologies can return results that have ecological validity relative to real-world use. The chapter introduces the complexities inherent in the mobile usage context and provides an overview of studies conducted to compare lab and field studies. The authors describe a study they conducted to compare the results obtained via lab and field testing. They recommend that, for most testing, it is best to perform several quick laboratory tests iteratively during the design process, rather than concentrate efforts on a single field test. They do, however, acknowledge that in some instances lab-based testing is insufficient—for example, due to technical limitations such as testing a GPS-based system, where it is difficult to simulate the use context with sufficient realism in the lab, or where it is the intention to observe user behaviour in a natural environment. In these situations the authors suggest field trials may be beneficial and, on the basis of their own experience, suggest some guidelines for conducting field tests of mobile technologies.

This chapter argues that, while mobile application evaluation protocols increasingly reflect user mobility, they fail to place realistic demands on users' visual attention (e.g., to reflect the real-life need for users to be cognizant of hazards as they move through their physical environment). In this chapter, the authors present a simple classification for describing the kind of distractions which might typically surround a user, and report on two evaluations designed to determine the effect visual distractions have on users of a mobile application. The results, which showed that users' requirement to monitor their environment affected both task performance and measures of workload, indicated that it is important to include such distractions along with mobility in evaluations of mobile technology.

This chapter introduces the concept of instrumented usability analysis for mobile devices – that is, the use of sensors (such as accelerometers) to elicit quantitative, objective information about the "moment

Section V
Case Studies

This chapter introduces the notion of a "closed environment", such as a hospital or military context, for which designers of technology often have extremely limited access to end users, both for design and testing purposes. During the design and development of such systems, many of the typical protocols employed in user-centered design are inapplicable; requirements gathering becomes an indirect process and quasi-lab studies are used in place of real contexts. This chapter reports, in detail, on a case study of the design of a mobile application to monitor vital signs of hospital patients. It discusses the analysis and design challenges faced, as well as the alternative evaluation methods that had to be devised to ensure ecological validity of evaluation results despite lack of direct access to users.

This chapter introduces the personal audio loop (PAL), an application designed to recover audio content from the recent past using the mobile phone platform. The authors discuss an evaluation of its potential usefulness in everyday life, the level of ubiquity and usability demanded of the service, and the social and legal considerations for long-term adoption. A detailed discussion of the various evaluation methods used (ranging from a controlled lab study to deployment of the system over a period of several weeks) is presented, as are analyses of the results obtained leading to an identification of issues critical to the use of PAL. This chapter raises and discusses interesting issues regarding the legality of a system such as PAL, highlighting that traditional privacy guidelines and policies may not adequately address personal ubicomp applications of this nature.

Mobile technology has the potential to provide convenient access to meeting records for users on the move. This chapter discusses issues surrounding the design, implementation, and evaluation of such interfaces, and proposes a general paradigm for meeting browsing which addresses the core information access requirements of the task within the constraints imposed by mobile technology. The authors illustrate their discussion with a case study and lessons learned developing a handheld meeting browser application.

This chapter describes the design of a collaborative m-Learning application that uses pair communication based on speech and text I/O. The authors present the process they undertook to develop their tool as a

Foreword

Today there are over two billion mobile phones in use. Add this vast number to the growing portfolio of mobile devices—from music and video players to portable and wearable medical and health monitors, from tiny tags to intelligent garments—and its easy to appreciate the importance and timeliness of a book on effective interface design and evaluation.

Poor interface design of these devices is at best a cause of frustration—*how do I zoom out so I can view that web page?* —at worse, life-threatening—*did the nurse enter that drug dosage correctly in my portable medical pump?* Meanwhile, with many billions of interactions occurring daily with mobiles, great design can transform the world for good: simply reducing the number of key presses by one or two on a popular service could save a lifetime of human effort. While good design benefits all of us, for users with physical and cognitive impairments, as chapters in this book illustrate, the potential positive impact is huge. The Handbook is an important tool to help us all as we strive for even better mobile interactions in the future.

There are, of course, many other books on human-computer interaction design and evaluation. Should you read this one? Answer: absolutely. Other texts can help you understand broad issues and approaches but mobiles are very different from desktop-bound conventional computers. Consider traditional HCI as being a visit to a metropolitan zoo; in contrast, mobile HCI is like doing an adventurous animal safari, deep in the bush.

Take just two aspects: contexts and tasks. Mobiles are used in highly dynamic and demanding environments: people want to look up directions while walking; check flight changes while encumbered with luggage, children and jet-lag; tell each other stories in cafes using photos and video stored on their gadget. The office where this is written is a much more predictable, calmer context! Then, while a lot of HCI research has tackled work-based systems and tasks, mobiles are forcing the community to tackle other user ‘goals’ like fun, curiosity, and connecting.

The book comprehensively probes the unique problems and opportunities facing mobile designers and researchers, giving a much richer picture than the other general HCI volumes. You will learn not only of technologies and designs but, importantly, of methods and tools you can use when you are engaged in building your own systems.

The mobile research and practice community has grown considerably over the past 10 years and the full spectrum of approach and focus is represented in the book. So, there are chapters concerned with ethnographic methods and accounts of mobile use; and, others which detail engineering innovations and experimentation. You’ll be able to learn about the state-of-the-art in interface and interaction technologies, from touch-based devices to RFID tagging; and be challenged by articles touching on issues like privacy, swarm intelligence, and technology acceptance models.

Whatever your area of mobile interest, you should be able to find material here that will inspire and inform design and evaluation processes. If you are a researcher, perhaps starting out on a particular topic such as mobile text entry, the book will provide you with good reviews of existing approaches and pointers to future challenges. If you are one of the many new developers working on designing mobile devices and services, the book also contains practical guidelines and case-study experience reports to help you make good choices. The book is not just for technologists, though: marketing and business strategists will be interested in the consumer and social analyses.

Mobiles are changing the world. People like you—researchers and developers—have an incredible opportunity to shape the future. This book will be a resource you should return to again and again to check and challenge your methods, tools, and approaches so that this future is enriched and not impoverished.

Matt Jones
Future Interaction Technology Lab
University of Wales Swansea
www.fitlab.eu

Matt Jones is a Senior Lecturer and is helping to set up the Future Interaction Technology Lab at Swansea University. He has worked on mobile interaction issues for the past ten years and has published a large number of articles in this area. He is the co-author of "Mobile Interaction Design", John Wiley & Sons (2006). He has had many collaborations and interactions with handset and service developers including Orange, Reuters, BT Cellnet, Nokia and Adaptive Info; and has one mobile patent pending. He is an editor of the International Journal of Personal and Ubiquitous Computing and on the steering committee for the Mobile Human Computer Interaction conference series. Married with three mobile, small children; when he's not working he enjoys moving quickly on a bike whilst listening to music and the occasional podcast.

Preface

In recent years, mobile technology has been one of the major growth areas in computing. Mobile devices are becoming increasingly diverse, and are continuing to shrink in size and weight. Although this increases the portability of such devices, their usability tends to suffer. Ultimately, the usability of mobile technologies will determine their future success in terms of end-user acceptance and, thereafter, adoption. Widespread acceptance will not, however, be achieved if users' interaction with mobile technology amounts to a negative experience. Mobile user interfaces need to be designed to meet the functional and sensory needs of users. In recognition of this need, a growing research area focusing on mobile human-computer interaction has emerged, and will likely continue to grow exponentially in the future.

The resource disparity between mobile and desktop technologies means that successful desktop user interface design does not automatically equate to successful mobile user interface design. Desktop user interface design originates from the fact that users are stationary (that is, seated at a desk) and can devote all or most of their attentional resources to the application with which they are interacting. As a result, the interfaces to desktop-based applications are typically very graphical (often very detailed) and use the standard keyboard and mouse to facilitate interaction. This has proven to be a very successful paradigm that has been enhanced by the availability of ever more sophisticated and increasingly larger displays. In contrast, users of mobile devices are typically in motion when using their device, which means that they cannot devote all of their attentional resources, especially visual resources, to the application with which they are interacting; such resources must remain with their primary task, often for safety reasons. Additionally, the form factor of mobile devices typically limits the applicability of standard input and output techniques, making mobile human-computer interaction design ineffective if we insist on adhering to the tried-and-tested desktop paradigm.

The design and evaluation of mobile human-computer interaction, unlike desktop-based interaction, needs to be cognizant of the implications brought to bear by complex contextual factors affecting both users and technology. Such contextual influences include, but are not limited to, the physical environment in which a mobile device is being used, the impact of multitasking behavior typically exhibited by users of mobile devices (e.g., using a device whilst driving), and the social context in which a device is used (e.g., consider social acceptability of interaction). All in all, designing the user interface for mobile applications is a very complex undertaking that is made even more challenging by the rapid technological developments in mobile hardware.

Not only is the design of human-computer interaction for mobile technologies difficult, so too is the evaluation of such designs. In fact, the most appropriate means by which to effectively evaluate mobile applications is currently a hotly debated topic in the field of mobile human-computer interaction. Evaluation techniques for mobile technology require as much consideration as the design of the user interfaces themselves; for the results of evaluations of mobile applications to be meaningful, the manner in which the evaluations are conducted needs to be, and is, the focus of considerable research in itself.

The purpose of the *Handbook of Research on User Interface Design and Evaluation for Mobile*

Technology is to offer a compendium of current research knowledge concerning the key issues surrounding the design and evaluation of mobile user interfaces such that students, researchers, educators, and practitioners alike may all derive benefit from the experience of leading experts working in this field. Its aim is to expose readers to, and heighten their awareness of, the complexity of issues concerning mobile human-computer interaction. Amongst the chapters included in the handbook, alternative points of view are included for some of the field's hotly debated topics in order to encourage readers to think out of the box and embrace the challenge of new paradigms both for interaction design and evaluation. Reliance on the tried-and-tested desktop design and evaluation paradigms has not worked; the mission of this handbook is to encourage people to *think out of the box* to ensure that novel, effective user interface design and evaluation strategies continue to emerge and, in turn, the true potential of mobile technology is realized.

To elicit the best and most balanced coverage of issues critical to the design and evaluation of mobile technologies, researchers from around the world were invited to submit proposals describing their intended contribution to the handbook. All proposals were carefully reviewed by the editor, with a view to assembling the finest contributions from leading experts in the field. Upon receipt of full chapter submissions, each submission was subjected to double-blind peer review, and only the best were then selected for final inclusion in the handbook. In many instances, the chapters were subjected to multiple revisions before final acceptance. The result of this rigorous process is a comprehensive collection of current research articles of high scholarly value written by distinguished researchers from many prominent research institutions and groups around the world.

ORGANIZATION OF THIS HANDBOOK

The goal of the *Handbook of Research on User Interface Design and Evaluation for Mobile Technology* is to improve our appreciation of the current and future challenges associated with the design and evaluation of user interfaces to mobile technologies. To achieve this goal, the handbook includes a comprehensive collection of 64 quality research contributions from leading experts around the world. It covers issues ranging from the use of ethnographic methods for design of mobile applications to instrumented lab-based methods for their evaluation. Additionally, each chapter includes a collection of related key terms and their definitions, contributing to a comprehensive compendium of terms, definitions, and concepts central to the field of mobile human-computer interaction.

Although most chapters touch on a number of the issues critical to user interface development for mobile technologies, and many include discussion of case studies for illustrative purposes, to assist you when searching for specific information, the 64 chapters have been organized according to their primary contribution. Hence, the handbook is organized into five sections that examine the following topics:

- **Section I: User Interface Design for Mobile Technologies**
  - Use of ethnography to inform mobile user interface design
  - Use of the technology acceptance mobile for mobile services to guide the design of mobile technologies
  - The impact of user characteristics on the design of mobile user interfaces
  - Wearable technologies and their design implications
  - Contextual information and awareness in mobile application design
  - Design of in-car user interfaces
  - Design of mobile learning applications
  - Adaptive and intelligent user interfaces in mobile computing

  - Rapid prototyping, modeling, and simulation tools for mobile applications
  - Ecologies of interacting artifacts for ubiquitous technologies
- **Section II: Novel Interaction Techniques for Mobile Technologies**
  - Classification of mobile interaction techniques
  - Novel interaction paradigms
  - Unobtrusive interaction
  - Visual interaction
- **Section III: Assistive Mobile Technologies**
  - Overview of key issues and trends for designing and evaluating mobile assistive technologies
  - Design for various special needs groups, including seniors, mental health interventions, and visually impaired users
  - Implications for designing the interface to smart wheelchairs
- **Section IV: Evaluation Techniques for Mobile Technologies**
  - Theoretical overview
  - Adaptation of traditional methods to suit mobile human-computer interaction
  - Method selection and combination strategies
  - Novel evaluation methods
  - Classification of usability factors for mobile technologies
  - Lab v. field evaluations
- **Section V: Case Studies**

The handbook provides literally thousands of references to existing literature and research efforts in the field of mobile human-computer interaction, and it includes a comprehensive index to support quick and convenient look up of topics and concepts. This handbook is an ideal reference for veteran and novice educators, researchers, students, and practitioners in the field of mobile human-computer interaction who require access to current information in this emerging field. The complementary combination of theoretical and practical content will enable readers to draw parallels with their own research or work, and apply and/or further the research efforts of others in their own projects.

Jo Lumsden
National Research Council of Canada
Institute for Information Technology – e-Business
Fredericton, Canada
2007

Joanna Lumsden is a Research Officer with the National Research Council of Canada's (NRC) Institute for Information Technology. Prior to joining the NRC, Lumsden worked as a research assistant in the Computing Science Department at the University of Glasgow, U.K. where she attained both her undergraduate software engineering Honours Degree and her Ph.D. in human computer interaction. Lumsden is also an Adjunct Professor at the University of New Brunswick in Fredericton, where she teaches graduate courses and supervises a number of graduate students. Lumsden is the lab manager for the NRC's Mobile Human Computer Interaction Lab – a facility dedicated to investigating mobile interaction design and evaluation.

Acknowledgment

Editing this handbook has been a unique, enlightening, and ultimately rewarding experience. I have learned a lot both personally and professionally, and have been amazed by the encouragement and enthusiasm I have received from the authors of this publication. This handbook has been 2 years in the making, and I would like to take a moment to acknowledge the efforts of those people who have made it possible.

My deepest gratitude goes to all the authors for their excellent contributions. I learned a lot from working with such knowledgeable and expert individuals around the world, and greatly appreciated their cooperation and kind words of encouragement throughout the process. Thank you! I hope the handbook does you proud!

Each chapter in this handbook was peer reviewed. I would like to acknowledge the assistance of *everyone* involved in the review process; their constructive reviews were invaluable to the quality of the handbook. In particular, I would like to thank two of my knowledgeable and insightful colleagues who stepped in and shouldered a heavier review load to help out when others let me down; Murray and Danny, thank you *so* much! Whilst I hope I have included all reviewers in the list, I appreciate there may be individuals of whom I am unaware who helped their colleagues review submissions "behind the scenes"; to any and all to whom this applies, my thanks.

I would like to extend my deep gratitude to the members of my Editorial Advisory Board: Dr. Gary Burnett from the University of Nottingham (UK), Dr. Murray Crease from the National Research Council of Canada (Canada), Dr. Mark Dunlop from the University of Strathclyde (UK), Mr. Philip Gray from the University of Glasgow (UK), Dr. Jesper Kjeldskov from Aalborg University (Denmark), Dr. Irina Kondratova from the National Research Council of Canada (Canada), Dr. Rod Murray-Smith from the University of Glasgow (UK), Dr. Antti Pirhonen from the University of Jyväskylä (Finland), and Dr. Janet Read from the University of Central Lancashire (UK). It was an honour to be able to turn to such a distinguished and wise group of people for support and sound advice during the collation of the handbook.

I am deeply indebted to the National Research Council of Canada's Institute for Information Technology for supporting me in this endeavour and sharing my belief in the value of this publication.

I owe a debt of gratitude to IGI Global for giving me the opportunity to realise my vision. In particular, thank you to Kristin Roth, my development editor, who guided me through the lengthy and complex process of publishing a book of this nature, and answered my myriad of questions and gave sound advice throughout.

I would like to thank Dr. Matt Jones from the Future Interaction Technology Lab at Swansea University (UK) for so willingly and enthusiastically writing the foreword to this handbook. Finally, my warmest thanks go to my husband, Keith, for his support, encouragement, patience, and love.

In closing, I am delighted to present this handbook and its comprehensive range of topics. I am certain that you will find it a useful reference, and source of better understanding, for all pertinent issues related to current research on the design and evaluation of mobile technologies.

Jo Lumsden
National Research Council of Canada
Institute for Information Technology – e-Business
Fredericton, Canada
2007

Section I
User Interface Design for Mobile Technologies

This section looks at many of the critical aspects concerned with effective design of mobile applications. The section begins with a series of chapters that discuss the adoption of ethnographic methods to inform the design of such technologies, including a selection of chapters that report on observed mobile device use and subsequent implications for design. This section covers issues such as how factors of user acceptance of mobile services can be used to guide the design of such technologies, as well as the impact of age and cognitive capacity on design. Chapters consider wearable technologies, the importance of contextual information in mobile application design, the design of in-car user interfaces, and issues surrounding the design and implementation of mobile learning applications. The section takes a look at adaptive and intelligent user interfaces for mobile computing, as well as tools for rapid prototyping, modeling, and simulation of mobile systems. The section concludes with a look to the future in terms of ecologies of interacting artifacts, reflecting an evolution from strictly mobile to more ubiquitous technologies.

Chapter I
From Ethnography to Interface Design

Jeni Paay
Aalborg University, Denmark

ABSTRACT

This chapter proposes a way of informing creative design of mobile information systems by acknowledging the value of ethnography in HCI and tackling the challenge of transferring that knowledge to interface design. The proposed approach bridges the gap between ethnography and interface design by introducing the activities of field-data informed design sketching, on a high level of abstraction, followed by iterative development of paper-based mock-ups. The outcomes of these two activities can then be used as a starting point for iterative prototype development—in paper or in code. This is particularly useful in situations where mobile HCI designers are faced with challenges of innovation rather than solving well-defined problems and where design must facilitate future rather than current practice. The use of this approach is illustrated through a design case study of a context-aware mobile information system facilitating people socialising in the city.

INTRODUCTION

This chapter looks at the mobile technology design problem of taking an ethnographic-based approach to gathering field data and making this data available to the design process in a form that is easily assimilated by designers to inform user-centred design of mobile technology. Interface design for mobile technologies presents unique and difficult challenges that sometimes render traditional systems design methods inadequate. Ethnography is particularly well-suited to design for mobile technology. Mobile usability is often highly contextual and ethnographic approaches can facilitate richer understandings of mobile use contexts providing insight into the user's perspective of the world. Exploring the huge potential of mobile devices presents designers with a unique opportunity for creativity. In thinking about mobile technology design for *future,* rather than *current* practice, the challenge becomes even greater.

Before this discussion proceeds further it is worth clarifying the use of the term *ethnography*. Traditionally, ethnographic studies within sociology are conducted from a particular theoretical viewpoint and for the purpose of contributing to theory. However, ethnography, as it is understood in HCI research, generally refers to a collection of techniques used for gathering and organizing field materials from observational studies (Dourish, 2006). By its very definition, ethnography is primarily a form of reportage. It provides both empirical observational data, and makes an analytical contribution in the organization of that data. The virtue of ethnography is that it takes place in real-world settings and provides access to the ways people perceive, understand, and do things (Hughes et al., 1997). Ethnographically-oriented field methods can be used in HCI to provide a deeper understanding of an application domain, a holistic understanding of users, their work, and their context, which can then be drawn into the design process at the earliest stages (Millen, 2000). Ethnographic studies involve detailed observations of activities within their natural setting, providing rich descriptions of people, environments and interactions, and acknowledging the situated character of technology use (Millen, 2000). These observations can provide valuable insights into the processes needed for systems requirements specifications (Sommerville et al., 1993).

In the literature, the terms ethnography and ethnomethodology are both used to refer to field studies using ethnographic methods to understand how people perceive their social worlds. Other terms such as technomethodology (Button & Dourish, 1996), rapid ethnography (Millen, 2000) and design ethnography (Diggins & Tolmie, 2003) are also used to distinguish different aspects of the use of ethnography in the design of technology. For the sake of simplicity, this chapter uses the term ethnography to encompass these understandings as being important to the discussion of the relationship between their outputs and the inputs they provide to the design process.

For ethnography to make a worthwhile contribution to the design of mobile technologies, we need to find ways for translating ethnographic findings into forms that are suitable for informing design processes. In the following sections, the historical relationship between ethnography and HCI is discussed, including how it has been incorporated into the process of interface design. The theoretical and methodological background for how to gather and interpret ethnographic data and use this for informing design is described. A design case study is then presented in which an ethnographic approach has been applied to mobile technology design in a real world research project through a structured series of activities. The overall process is described, and the two steps of developing *design sketches* and *paper-based mock-ups* are introduced as a way of bridging the gap between ethnography and interface design. Finally, lessons learned from using design sketches and paper-based mock-ups in the development process are outlined.

BACKGROUND

Ethnography and HCI

The issue of bridging the gap between ethnography and interface design has been a topic of discussion in HCI research for over a decade. Ethnography is now regarded as a common approach to HCI research and design (Dourish, 2006). Yet there is still no overall consensus on how best to incorporate the results of ethnographic fieldwork into the design processes (Diggins & Tolmie, 2003). In the early 90s seminal work by sociologists, such as Suchman, Hughes, Harper, Heath and Luff, inspired the use of ethnography for understanding the social aspects of work processes and informing user interface design (Hughes et al., 1995). However, researchers struggled with the challenge of utilizing insights provided by ethnography into the activity of designing. By the mid 90s, ethnography was hailed as a new approach to requirements elicitation for interactive system design, particularly through its application in the development of computer-supported cooperative work (CSCW) systems (Hughes et al., 1995). Even so, some researchers still held reservations about the ability of ethnographic methods to inform design (Hughes et al., 1997) and ethnography was regarded as a relatively untried approach to systems development, despite the fact that it was increasingly being used to inform and critique actual systems (Button & Dourish, 1996). Toward the end of the 90s, researchers were beginning to develop systematic approaches to social analyses for the purpose of influencing design (e.g., Viller & Sommerville, 1999). However, despite many research efforts, bridging the gap between ethnography and design still remains a matter of concern to HCI researchers today (Diggins & Tolmie, 2003).

The turn towards ethnography within HCI was motivated by a growing need to design for complex real world situations. This began with the belief that methods from the social sciences, such as ethnography, could provide means for understanding these contextual issues of technology use better. In the light of today's ubiquitous and mobile networked computing environments, the need to understand contexts of technology use, such as peoples' dynamic work and social practices, is challenging HCI researchers and designers more than ever. Supporting innovation in a world of emerging technologies can be done by submerging designers, who understand emerging technical possibilities, into rich ethnographic field data about potential users' lives and current practices (Holtzblatt, 2005). In this way technology design drives an understanding of the user's situation, which in turn, propels innovation.

Ethnography and Interface Design

The process of transition from field data to prototype design is a difficult one (Cheverst et al., 2005; Ciolfi & Bannon, 2003). A design process involving ethnography generally starts with observations and interviews collected through ethnographic methods. Key findings are then summarized and design ideas are drawn out with a set of features that can be tied back to the findings. The next step involves, "design suggestions" or "design implications," which may evolve into requirements through the development of a low-fidelity prototype. This prototype is then iterated with feedback from users and evolves into the operational system. The data collected by ethnographic methods reflects the richness of the user's situation in a way that is difficult to derive from a limited set of questions or measures as employed in traditional analysis methods (Wixon, 1995). In contrast to traditional systems analysis that looks at data, structures, and processing, ethnography is concerned with participants and interactions (Sommerville et al., 1993). This provides the designer with a rich understanding of the context of use for the artifacts that are being designed (Millen, 2000). In looking at a situation through the user's eyes rather than the designers, ethnography provides a view of the situation that is independent of design preconceptions (Hughes et al., 1997).

Ethnography has much to contribute to interface design—particularly in mobile device design due to the highly contextual nature of mobile usability and use. However, one of the main problems is finding a suitable mechanism for the transference of knowledge between these two fundamentally different disciplines. Ethnographic findings need to be understood and communicated to designers (Hughes et al., 1995). And yet, current mechanisms for incorporating ethnographic findings into the design process still fail to capture the value of these investigations (Dourish, 2006).

Ethnography deals in "the particular," and software design in "the abstract" (Viller & Sommerville, 1999). While willing to listen to each other, both disciplines speak different languages and use different methodologies. Ethnographers deal in text, notes, reports, and transcriptions, and produce detailed results giving a rich and concrete portrayal of the particulars of everyday practical action in context, presented in a discursive form; software designers and engineers deal in the creation and manipulation of more formal graphical abstractions, notations and description techniques to simplify the complexity of the situation and extract critical features. Ethnographers avoid judgements; designers make them. Where ethnographers take an analytic role, including gathering and interpreting data, software designers have a synthesis role, designing from abstract models of situations (Button & Dourish, 1996; Hughes et al., 1995). In addition to the problems of communication there are also problems of timing. Ethnography is generally conducted over a long period of time; in fact, it is difficult to define an end point for gathering understanding. On the other hand, software designers are often under restricted time pressure to deliver a product.

The problem has been in finding a timely method and a suitable form to present field findings that can be assimilated by and are readily usable for designers (Hughes et al., 1995; Viller & Sommerville, 1999). The needs of the software designer have to be aligned with a representation of the essential "real world" practices of users in context. Simply describing the social events being observed is not sufficient, designers need to be able to model and use this understanding in design.

USING ETHNOGRAPHY IN THE DESIGN PROCESS

Gathering Data

From HCI research it can be seen that using ethnography as a data gathering method requires the development of more structured approaches to conducting and reporting from ethnographic studies that better support the development of design requirements.

One approach is to conduct ethnography concurrently with design and bring ethnographic results into the design process in a more systematic way throughout the development process. This can, for example, be achieved through meetings between ethnographers and the design team (Hughes et al., 1995). This approach results in a change in the way that ethnography is conducted. Rather than extended periods in the field, ethnographers working in cooperation with software designers to create a system design, making short and focused field studies, reporting back to designers, and often taking design questions back into the field to focus their observations and questions to users. To structure the process, the communication of fieldwork to designers can be supported by dedicated software packages (Diggins & Tolmie, 2003; Sommerville et al., 1993). In this situation, the ethnographic record becomes a joint resource with ethnographers regularly reporting their findings in an electronic form, and designers using this content to develop structured design requirements. Constructing these records in a connected manner preserves backward and forward traceability between ethnographic findings and evolving system requirements.

Another approach is to lead into the design process through *rapid ethnography* (Millen, 2000). Rapid ethnography provides the field worker with a broad understanding of the situation which can then be used to sensitize designers to the use situation rather than identifying specific design issues. It is aimed at gaining a reasonable understanding of users and their activities in the short time available for this in a software development process. Rapid ethnography provides a more structured approach to ethnographic field studies by limiting the scope of the research focus before entering the field. It focuses time spent in the field by using key informants in the real situation and interactive observation techniques. Rapid ethnography also uses multiple observers in the field to ensure several views of the same events and to create a richer representation and understanding of the situation (Millen, 2000).

Interpreting Data

Ethnography is not simply about the collection of data in the field, it is also about reflection on and interpretation of that field data. Effective communication between ethnography and design is at the heart of the matter of bridging the gap between the two disciplines (Hughes et al., 1997). By recognizing the different natures and input and output requirements of ethnography and interface design, integration between the two disciplines can be achieved through enhancing and structuring the communication between them during the interpretation phase.

One approach to interpreting the data collected is to have a cross-discipline team participating in the fieldwork. In this situation designers go into the field with ethnographers to experience themselves how users work. They also contribute to the representation of the gathered data, shaping it into a form that is easier for designers to use (Diggins & Tolmie, 2003). Representing ethnographic findings through pictorial stories, drawings, data models, analogies and metaphors are ways to communicate field learning to cross-discipline teams (Millen, 2000). Videotapes of field observations and design documentaries play a similar role using a more designer-accessible communication mode than a written report (Raijmakers et al., 2006).

Another approach to interpretation is to have both ethnographers and designers involved in the conceptual design process. In this situation, the ethnographer is an ongoing member of the design team, providing grounded insights and interpretations into the abstracted requirements as they evolve and the design emerges. The ethnographer acts as a substitute user during the design process (Viller & Sommerville, 1999). Through their knowledge of the actual situation, they can participate in discussions with the designers, providing insights and access to instances of specific relevant situations.

A third approach is for the designer to play the part of a pseudo-ethnographer. This involves designers going "into the wild" and being exposed to users by watching real work while it is being done, and hence truly experiencing the richness of work (Wixon, 1995). Structured methods such as rapid ethnography and *contextual design* (Beyer & Holtzblatt, 1998) make this possible. In contextual design, the user and the designer explore the design space together using *contextual interview* or *facilitated enactment* of their practices in context (Holtzblatt, 2005). *Affinity diagramming*, from the contextual design method, provides a synthesis of the data into hierarchical classifications where the meaning contained in the data elements can be reflected on in relation to the design question, facilitating understanding and innovation for designers.

Informing Design

After the ethnographically gathered field data has been interpreted, abstracted findings are used to derive design opportunities and design requirements. The designer uses the outputs from the interpretation of the field data as input into the design process. Sometimes the ethnographers are involved in this design process bringing their intimate knowledge of the users and the situation of use, and their deep relationship to the data, to the team (Cheverst et al., 2005). They participate in the identification of design incentives by drawing attention to general design opportunities, and relevant topics and concerns. Otherwise, the designers must draw understanding entirely from the reports, discussions, diagrams and models, which represent the ethnographic record.

Design is a matter of making, and is used to create and give form to new ideas and new things (Fallman, 2003). A recent approach to informing design and achieving a close connection between the design team and the field data is the use of field observation videos or design documentaries. These videos mediate between ethnographic and design perspectives. As the design team watches them they incorporate interpretation of data into the design process on the fly through discussions drawing design sensitivities and identifying design concerns. Designers become sensitized to relevant issues visible in the real world interactions depicted in the video (e.g., Ciolfi & Bannon, 2003; Raijmakers et al., 2006). This method requires a high level of design experience, and in bridging the gap between ethnography and design, these designers work in an inspirational, ephemeral and creative way. For others this creative leap across the divide is very difficult, and more structured methods are needed to guide the process of envisioning design from ethnographic outputs. In response to new interface design challenges, including mobile technology, HCI researchers are investigating new techniques for guiding designers through this difficult transition – of particular interest to this chapter are the techniques of *design sketching* (Buxton, 2007), *paper-based mock-ups* (Ehn & Kyng, 1991) and *paper prototyping* (Snyder, 2003).

Design sketching is fundamental to the process of design, and can be used by information system designers to bring about the realization of an idea in the way designers think (Fallman, 2003). Sketching is the art of giving form to the unknown; it makes it possible to "see" ideas or envision whole new systems, and is especially critical in the early ideation phase of design (Buxton, 2007). According to Buxton, sketches should be rapid, timely, inexpensive, disposable, plentiful, clear, un-detailed, light, informal representations that practitioners can produce and interact with to suggest and explore ideas. Sketching is not only a way to visualize existing ideas, but it is about shaping new ideas. In making a sketch of something, the visualization talks back to the designer with a new perspective on that idea, providing a link between vision and realization of new ideas.

Paper-based mock-ups are closely related to the notion of design sketching. In this technique from the participatory design tradition, representational artifacts are constructed from paper, cardboard and materials at hand. Informed by studies of practice, mock-ups can play an important mediating role in connecting use requirements and design possibilities in a form recognizable to multi-disciplinary design teams (Ehn & Kyng, 1991). These mock-ups can be used to incorporate materials from the ethnographic study, embody envisioned new technological possibilities, convey design ideas in relation to existing practices and reveal requirements for new practices (Blomberg & Burrell, 2003).

Paper prototyping is a widely used technique for designing, testing and refining user interfaces (Snyder, 2003). This technique helps with the development of interfaces that are useful, intuitive, and efficient, by initiating testing of the interface at a stage when the design is in its formative stages and therefore still open to the input of new ideas. Paper prototyping can be used to reflect on field study findings while developing and refining the design (Holtzblatt, 2005). A collection of interface designs, drawn from ideas generated through design sketching and paper-based mock-ups are given functional and navigational connections through the process of paper prototyping. A paper prototype is a useful vehicle for giving visual form to identified design requirements. It forms the focus for design refinement discussions and cognitive walkthroughs by the design team, and is in itself part of the design specification for implementation of the system.

A DESIGN CASE STUDY

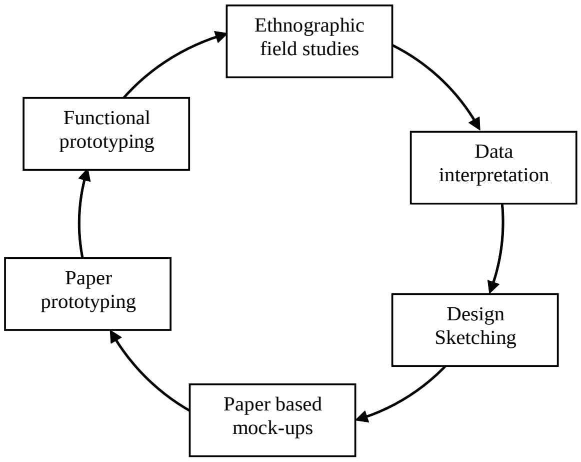
The project used as a design case study in this chapter involved the development of a context-aware mobile information system, *Just-for-Us*, designed to facilitate people socialising in the city by providing information about people, places, and activities in the user's immediate surroundings. The case study location was a specific city precinct covering an entire city block, Federation Square, Melbourne, Australia. This location was chosen because it is a new, award-winning architectural space providing a variety of activities through restaurants, cafes, bars, a museum, art galleries, cinemas, retail shops, and several public forums spanning an entire city block. The design intention for the civic space was to incorporate digital technologies into the building fabric creating a combination of virtual information space and physical building space for people to experience. Thus, this particular place provided a unique setting for studying people's situated social interactions in a "hybrid" space and for inquiring into the user experience of mobile technology designed to augment such a physical space with a digital layer.

Process

The Just-for-Us mobile information system was designed specifically for Federation Square on the basis of an ethnographic study of people socialising there. The development process involved seven major activities:

- Ethnographic field studies
- Field data interpretation
- Design sketching on a high level of abstraction

Figure 1. The overall process of designing the Just-for-Us mobile information system



- Paper-based mock-up development
- Iterative paper prototyping
- Implementation of a functional prototype
- Field studies of prototype use in-situ

The specific content and outcome of these activities are described in the following subsections. Details of the implemented system and findings from the field study of its use are not covered here, but can be found in Kjeldskov and Paay (2006).

As illustrated in Figure 1, data from ethnographic field studies of situated social interactions in public were subjected to data interpretation, using the *grounded theory* approach (Strauss & Corbin, 1990) and affinity diagramming (Beyer & Holtzblatt, 1998). In trying to bridge the gap between our ethnographic data and actual mobile device interface design, outcomes from the interpretation of field data were used to inform a systematic activity of design sketching (Buxton, 2007). The purpose of this activity was to generate design ideas on a high level of abstraction inspired by ethnographic findings but without getting into too much detail about specific look, feel and functionality. On the basis of selected design sketches, we developed a number of paper-based mock-ups (Ehn & Kyng, 2991) of potential design solutions. This forced us to become more specific, but still allowed us to focus on overall functionality and interaction rather than on technical details. After this, we engaged in a number of paper prototyping (Snyder, 2003) iterations with the purpose of developing a detailed set of system requirements and a coherent interface concept prior to writing any program code. Finally, these specifications were implemented in a functional prototype allowing us to introduce new technology into the field and revisit peoples' socialising behavior in the city while using the operational Just-for-Us context-aware mobile information system.

Gathering and Interpreting Data

The aim of our ethnographic field study was to inquire into peoples' social interactions at Federation Square. The field study was guided by a subset of McCullough's typology of everyday situations (McCullough, 2004) for classifying peoples' social activities when out on the town: eating, drinking, talking, gathering, cruising, belonging, shopping, and attending. The study applied a rapid ethnography approach and consisted of a series of contextual interviews (Beyer & Holtzblatt, 1998) and ethnographic field observations (Blomberg & Burrell, 2003) with the designers acting as pseudo-ethnographers and gathering the field data (Figure 2). Three different established social groups participated in the study. Each group consisted of three young urban people, mixed gender, between the ages of 20 and 35, with a shared history of socialising at Federation Square. The groups determined the activities undertaken and the social interactions that they engaged in. Prior to the field visits, each group received a 10-minute introduction to the study followed by a 20-minute interview about their socialising experiences and preferences. This introduction occurred at a place familiar to the group, where they might meet before socialising in the city. This encouraged them to reflect on past social interactions, to relax about the visit, and gave the interviewer insight into the situated interactions that the group typically participated in. One of the members of the group was then taken to Federation Square and asked to arrange to meet up with the other members of the group. The group was then asked to do what

Figure 2. Ethnographic observations and contextual interviews at Federation Square

they would usually do as a group when socialising out on the town—while "thinking aloud" as they moved around the space, and responding to questions from the interviewer. Two researchers were present in the field, providing multiple views on the data collected.

Each field visits lasted approximately three hours and allowed the groups to engage in a number of social activities. The outcome of the ethnographic field studies amounted to eight hours of video and approximately 30 pages of written notes.

In addition to the observational studies of people socialising at Federation Square an architecturally trained observer carried out a single *expert audit* (Lynch, 1960) focusing on the physical space of Federation Square. The expert audit documented architectural elements and their relationships to surrounding context, including the people inhabiting the space through 124 digital photographs and corresponding field notes.

Interpreting data gathered from the ethnographic study involved two phases. Firstly, photographic data and written notes from the expert audit were analyzed using *content analysis* (Millen, 2000) and affinity diagramming (Beyer & Holtzblatt, 1998). Concepts and themes describing the physical space of Federation Square were overlaid onto a map of the precinct to produce a color-coded multi-layered abstraction of the space (Figure 3). This provided an overview of the spatial properties of Federation Square highlighting constraints and enablers for situated social interactions there with traceable links back to specific observations.

Secondly, video data from the contextual interview and observational field study of people socialising at Federation Square was transcribed and then analyzed using open and axial coding adapted from *grounded theory* analysis (Strauss & Corbin, 1990). Identifying key words or events in the transcript, and analyzing the underlying phenomenon created the initial open codes. Analysis of these codes resulted in a collection of categories relating to actions and interactions. After the codes were grouped into categories, higher-level themes were extracted using axial coding. Affinity diagramming was then used to draw successively higher levels of abstraction from the data by grouping and sorting the themes until a set of high-level concepts, representing the essence of the data and encompassing all lower level themes, had been formed. The process of affinity diagramming produced a hierarchical conceptual framework containing three overall clusters of themes abstracted from the transcripts (Figure 4). This provided a rich story about how people interact with each other while socialising in public, with traceable links back to specific observations in the field study sessions.

As illustrated in Figures 3 and 4, outcomes from the interpretation of our ethnographic field data were primarily on an abstract level, providing a deeper understanding of peoples' situated social interactions in the physical space of Federation Square. While this is an important part of the foundation for good design, in their current form these outputs did not point towards any particular design ideas. As an example, the analytical outcomes from interpreting the field data included a series of qualitative statements similar to those in the following list (For a detailed account of findings from the ethnographic field studies see Paay and Kjeldskov (2005)).

- Federation Square has four key districts with distinctly different characteristics, each with an associated landmark.

Figure 3. Graphical image of inhabited social context at Federation Square

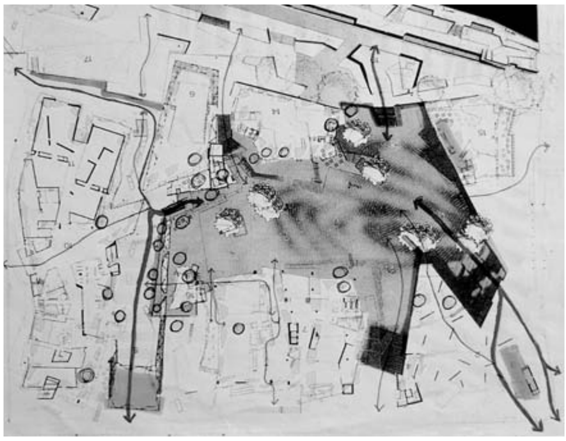
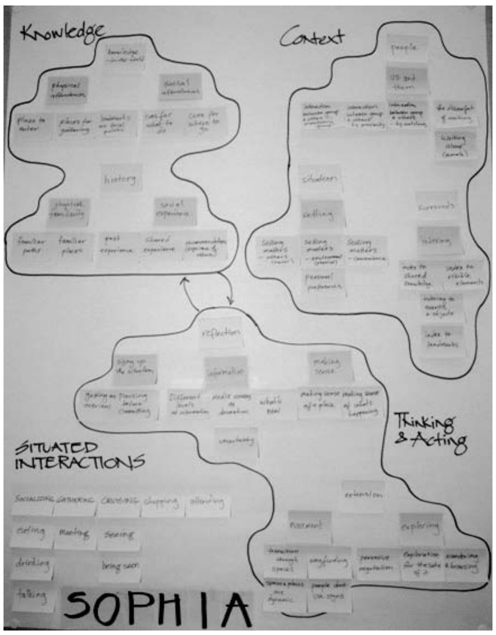


Figure 4. Affinity diagram of situated social interactions at Federation Square



- Federation Square has visible surroundings, general paths, general entrances, focal structures and no clear paths, so people need to use the structures and surrounds in finding their way around the space.
- People socialising at Federation Square like getting an overview of what is happening around them, and want to know about the presence and activities of other people.
- People's past experience with places and people at Federation Square play an important factor in choosing places and activities for socialising.
- People give directions at Federation Square by referring to shared experiences and visible elements, and use their history and physical familiarity with a place to find their way around using familiar paths.

In order to move forward from data interpretation toward an overall design concept as well as actual interface design and system requirements for a context-aware mobile information system for people socialising at Federation Square, the design team engaged in two steps of developing design sketches and paper-based mock-ups (as described earlier). Each of these techniques produced interface design artifacts on different levels of detail and abstraction. These two "bridging" steps between ethnography and interface design are described in the following sections.

Design Sketching

The first step in the design of the Just-for-Us mobile information system was to develop a series of conceptual design ideas based on the insight from our data analysis. For this purpose, the design team spent two days generating, discussing, sketching, and refining design ideas on the basis of the abstract models of the architectural space of Federation Square and the clustering of themes in the affinity diagram from the analysis of people socialising there.

The design sketching activity was done in a dedicated design workspace with sheets of A1 paper lining the walls on which we could sketch and refine design ideas. Each sketch took its origin in a specific finding or observation from the interpreted field data. This field finding would firstly be discussed in more detail to ensure shared understanding among the design team. Secondly, we would start sketching possible design ideas, for example, how to facilitate an observed practice.

Figure 5. Design sketching informed by interpreted ethnographic field data. The delineated area corresponds to the paper-based mock-up produced later and highlighted on Figure 6.

Hence, we were, in a sense, using collaborative data analysis, as described in the rapid ethnography method, to drive the generation of design ideas.

During the process of sketching, the conceptual outcomes from the data interpretation phase were continually revisited and, in turn, the sketches were continuously annotated with post-it notes referring to the data. For example, a section of the affinity diagram included the themes of "social experience," encompassing "past experience" and "shared experience." A diagram was then sketched to explore the intersections between past and shared experiences in groups of friends. In this way, we ensured a strong link between data and design, and maintained clear traceability between the two. This activity was about sketching the social concepts that came out of the data models, not about generating solutions. In doing this, we were able to explore the field data findings in a graphical form, and to explore derivations from these concepts by generating multiple understandings of them. Design sketching was used as a mechanism to understand the field outcomes, to generate graphical overviews of the design space, and create graphical representations of design opportunities within that space.

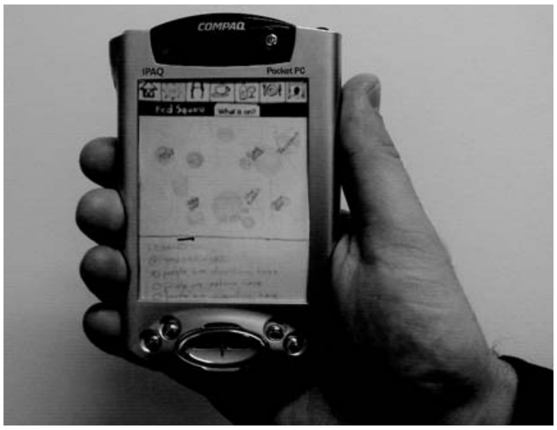
The outcome from the two-day design workshop was a collection of design sketches on A1 paper (Figure 5), each describing conceptually a potential design idea or design opportunity, for parts of the Just-for-Us mobile information system, including envisioned general functionality, general ideas for graphical design and user interaction, with clear references back to the empirical data.

The design sketches provided a new visual abstraction to the ethnographically interpreted field data, translating understanding encapsulated in the abstract findings into design parlance. Engaging in the process of design sketching rather than jumping straight to specifying system requirements, enabled us to see the ethnographic findings from a new perspective and to play with design ideas on a high level of abstraction. This allowed us to distance ourselves from the role of "problem solvers" and to explore instead, on a conceptual level, design ideas facilitating potential future practice in technology use.

Paper-Based Mock-Ups

While useful for generating and working with overall design ideas, conceptual design sketches are far too abstract for informing specific system requirements. Hence, moving directly on to detailed prototype design and implementation is likely to commit designers to specific solutions too early and impede their flexibility to try out new ideas. In an attempt to overcome this problem, the next step of our process from ethnography to interface design was to produce a series of

Figure 6. One of the paper-based mock-ups of possible mobile device screens



paper-based mock-ups of possible specific design solutions (Figure 6).

The production of paper-based mock-ups took place over several days and facilitated a series of long discussions within the design team leading to an overall concept for the Just-for-Us mobile information system providing functionality such as: an augmentation of the user's physical surroundings; chat capability with friends out on the town; content indexed to the user's physical and social context and history of interactions in the city; a graphical representation of places, people and activities within the user's vicinity; and way-finding information based on indexes to landmarks and familiar places. These design ideas were screen-based solutions to design opportunities identified during design sketching.

Working with each of these ideas in more detail, the paper-based mock-ups gave the design team a medium for trying out and modifying specific design ideas for what the system should be able to do and what it should look like—long before any actual coding was done. Consequently, the mock-ups coming out of this activity had already undergone several iterations of redesign and refinements.

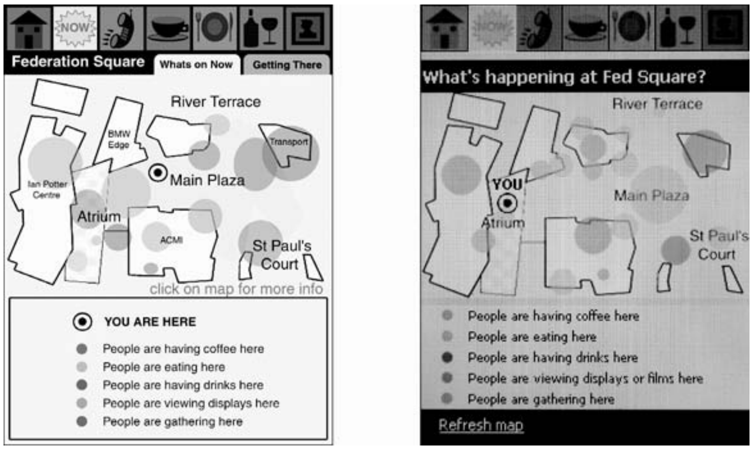
Discussions during the mock-up phase took place on different levels of abstraction: from screen design, system functionality, privacy issues, problems designing for small screens, what aspects of the user's context to capture in the system, and how to do this. We also had several discussions about whether or not the implementation of the produced mock-ups would be feasible within current mobile technologies, and if not, which enabling technologies would have to be developed. Through these discussions and continued refinements and redesigns, a set of specific design requirements slowly began to take shape—gradually taking us into the "safer ground" of interface design.

Prototyping

Having completed the paper-based mock-up phase, the final steps of our development process were much more straightforward. On the basis of the mock-ups, more detailed paper prototypes were produced using Adobe Photoshop (Figure 7 left). This forced the design team to work within the graphical limitations of the target device and to use the specific graphical user interface elements available in the target browser, for this web based application. Also, the detailed paper prototypes allowed the designers to discuss some of the more dynamic interaction issues such as navigation structure and handling of pushed information. While most design changes were done in Adobe Photoshop at this time, some of the more serious issues, such as how to fit the Internet chat screen(s) into the limited design space, forced the design team back to working with paper-based mock-ups for a short time. After several cognitive walkthroughs, a full paper prototype with a detailed set of requirements was agreed upon and implemented as an operational mobile web site providing context-aware information to users, with very few modifications (Figure 7 right).

The design specified by the paper prototype was implemented as a functional Web-based system accessible through the Web browser of a PDA (personal digital assistant) providing context-related information, dynamic maps and location specific annotated graphics to the user. It also keeps a history of the user's visits to places around the city. The functional prototype uses WLAN or GPRS for wireless Internet access and resolves the user's location and the presence of friends in vicinity by

Figure 7. Detailed paper prototype screen (left) and the corresponding final functional prototype screen (right), designed from the paper-based mock-up highlighted in Figure 6



means of Bluetooth beacons potentially embedded into the environment. The implementation of a functional prototype allowed us to close the circle depicted in Figure 1 by returning to Federation Square to do an ethnographic field study of people socialising there—this time facilitated by the Just-for-Us system. For details on this use study see Kjeldskov and Paay (2006).

FUTURE TRENDS

The future trends for bridging between ethnography and interface design for mobile technologies are many. As a part of a drive toward more user centered innovative design for both current and future practice, new techniques are emerging, which respond to the specific challenges of mobile technology design and use. These include, for example, cultural probes, digital ethnography, video diaries, film documentaries, facilitated enactment, acting-out in context, role-playing and body storming. Through these new techniques, the roles of ethnographers, designers, and future users are becoming more interwoven, facilitating a smoother and more effortless transition from ethnography to interface design. Techniques such as these reflect the fact that mobile technology design is not only about designing for existing work practices but also about designing for future practices in peoples' private and social lives and responding to the challenge of innovating for non-work in as yet non-existing use situations. They also respond to issues raised by many researchers that mobile technologies are often used in dynamic and continually changing contexts, offering information directly related to those contexts, and that it can be very difficult to predict what future user-adaptations of mobile technology might evolve.

The techniques of sketching and mocking-up introduced in this chapter are not new. Both have a long tradition in other design disciplines. However, like many of the above emerging approaches, we have combined existing techniques in a new way that provides designers with a more structured path to follow when making the difficult transition of transferring knowledge from the field into the design process.

CONCLUSION

This chapter addresses the issue of ethnography informing interface design for mobile technologies. It has described how ethnographic studies can be used in HCI design and how such studies can be useful for understanding current practice as well as providing a backdrop for envisioning potential future practice. However, as confirmed in the literature, bridging between ethnography and design is difficult, and techniques are needed that enable designers to better use ethnographic findings in the design process. In response to this, the two steps of conceptual design sketching and creating paper-based mock-ups have been proposed as bridging activities between ethnographic data interpretation and iterative prototype development.

Illustrating how this can be done in practice, this chapter has described a recent project involving the design of a context-aware mobile information system on the basis of a rapid ethnographic field study. In this project, the process of design sketching from analytical data made a useful link between interpretation and design. It provided a means of communicating a conceptual understanding of current practice into the early stages of interface design, and helped "translate" findings from the field data into design parlance. Working with sketches allowed the design team to play with design ideas on a conceptual level rather than moving straight to specifying system requirements. It also allowed them to distance themselves from the role of "problem solvers" and to explore instead potential future practice of technology use.

The process of creating and refining paper-based mock-ups on the basis of selected design sketches gave the design team a medium for being a bit more specific while still maintaining a high level of flexibility. It allowed for drilling down into some specific design ideas and the exploration and modification of ideas for interface design and functionality before doing any coding. It also allowed the team to engage in discussions about possible screen designs, different functionality, privacy, small screens, etc., and to rapidly implement, evaluate, and refine design ideas. By working with paper-based mock-ups, it was possible to generate a strong set of specific design requirements, which provided a solid foundation for subsequent activities of paper and functional prototyping.

Innovative interface design for mobile technologies is both an art and a science. It requires us to be creative and inspired as well as structured and focused. Facilitating creativity and inspiration provides the art. Grounding interface design in empirically informed understanding of people and current practice provides the science. The challenge we are faced with is not just how to perform the art and science of design better individually, but more so how to support a fruitful interplay between the two. For this purpose, techniques such as conceptual design sketching and creation of paper-based mock-ups are valuable tools for researchers and designers on their journey from ethnography to interface design.

REFERENCES

Beyer, H., & Holtzblatt, K. (1998). *Contextual design—Defining customer centred systems*. San Francisco: Morgan Kaufmann.

Blomberg, J., & Burrell, M. (2003). An ethnographic approach to design. In J. Jacko & A. Sears (Eds.), *Handbook of human-computer interaction* (pp. 964-986). Mahwah, NJ, USA: Lawrence Erlbaum Associates Inc.

Button, G., & Dourish, P. (1996). Technomethodology: Paradoxes and possibilities. In *Proceedings of CHI 96*, (pp. 19-26). Vancouver, Canada: ACM.

Buxton, B. (2007). *Sketching user experiences: Getting the design right and the right design*. San Francisco, Morgan Kaufman Publishers.

Cheverst, K., Gibbs, M., Graham, C., Randall, D., & Rouncefield, M. (2005). Fieldwork and interdisciplinary design. *Notes for tutorial at OZCHI 2005*. Retrieved October 24, 2007, from http://www.comp.lancs.ac.uk/rouncefi/Tutout.html

Ciolfi, L. & Bannon, L. (2003). Learning from museum visits: Shaping design sensitivities. In *Proceedings of HCI International 2003* (pp. 63-67). Crete, Greece: Lawrence Erlbaum.

Diggins, T., & Tolmie, P. (2003). The 'adequate' design of ethnographic outputs for practice: some explorations of the characteristics of design resources. *Personal and Ubiquitous Computing, 7*, 147-158.

Dourish, P. (2006). Implications for Design. In *Proceedings of CHI 2006* (pp. 541-550). Montreal, Canada: ACM.

Ehn, P., & Kyng, M. (1991). Cardboard computers: Mocking-it-up or hands-on the future. In J. Greenbaum & M. Kyng (Eds.), *Design at work: Cooperative design of computer systems* (pp. 167-195). Hillsdale, NJ, USA: Lawrence Erlbaum Associates, Publishers.

Fallman, D. (2003). Design-oriented human-computer interaction. In *Proceedings of CHI 2003* (pp. 225-232). Florida, USA: ACM.

Holtzblatt, K. (2005). Customer-centred design for mobile applications. *Personal and Ubiquitous Computing, 9*, 227-237.

Hughes, J., King, V., Rodden, T., & Andersen, H. (1995). The role of ethnography in interactive systems design. *Interactions, 2*(2), 56-65.

Hughes, J., O'Brien, J., Rodden, T., & Rouncefield, M. (1997). Designing with ethnography: A presentation framework for design. *In Proceedings of DIS '97* (pp. 147-158). Amsterdam, Holland: ACM.

Kjeldskov, J., & Paay, J. (2006). Public pervasive computing in the city: Making the invisible visible. *IEEE Computer, 39*(9), 30-35.

Lynch, K. (1960). *The image of the city*. Cambridge, MA, USA: The MIT Press.

McCullough, M. (2004). *Digital ground—Architecture, pervasive computing and environmental knowing*. Cambridge, MA, USA: The MIT Press.

Millen, D. R. (2000). Rapid ethnography: Time deepening strategies for HCI field research. In *Proceedings of DIS ´00* (pp. 280-286). Brooklyn, NY: ACM.

Paay, J., & Kjeldskov, J. (2005). Understanding situated social interactions in public places. In *Proceedings of Interact 2005* (pp. 496-509). Rome, Italy: Springer-Verlag.

Raijmakers, B., Gaver, W., & Bishay, J. (2006). Design documentaries: Inspiring design research through documentary film. In *Proceedings of DIS 2006* (pp. 229-238). Pennsylvania, USA: ACM.

Snyder, C. (2003). *Paper prototyping*. San Francisco: Morgan Kaufmann Publishers.

Sommerville, I., Rodden, T., Sawyer, P., Bentley, R., & Twidale, M. (1993). Integrating ethnography into the requirements engineering process. In *Proceedings of IEEE International Symposium on Requirements Engineering* (pp. 165-181). San Diego, CA, USA: IEEE Computer Society Press.

Strauss, A. L., & Corbin, J. (1990). *Basics of qualitative research*. Newbury Park, CA, USA: Sage Publications.

Viller, S., & Sommerville, I. (1999). Coherence: An approach to representing ethnographic analyses in systems design. *Human-Computer Interaction, 14*, 9-41.

Wixon, D. (1995). Qualitative research methods in design and development. *Interactions, 2*(4), 19-24.

KEY TERMS

Affinity Diagramming: One of the techniques of the contextual design process, used during data interpretation sessions to group related individual points together, creating a hierarchical diagram showing the scope of issues in the work domain being studied.

Content Analysis*:* A qualitative research technique for gathering and analyzing the content of text, where content can be words, meanings, pictures, symbols, ideas, themes, or any message that can be communicated, to reveal messages in the text that are difficult to see through casual observation.

Contextual Design: A collection of techniques supporting a customer-centered design process, created by Beyer and Holtzblatt (1998), for finding out how people work to guide designers to find the optimal redesign for work practices.

Design Sketch: A graphical representation of a concept or design idea on a high level of abstraction. It should be quick, timely, open, disposable, un-detailed, and informal, and is usually hand-drawn on paper.

Expert Audit: A field reconnaissance done by an architecturally trained observer maping the presence of various elements of the physical environment and making subjective categorizations based on the immediate appearance of these elements in the field and their visible contribution to the image of the city.

Ethnography: A collection of techniques used for gathering and organizing field materials from observational studies, involving detailed observations of activities within their natural setting, to providing rich descriptions of people, environments and interactions.

Grounded Theory: A theory based analytical approach, which takes a set of data collected using ethnographic methods and provides a set of specific procedures for generating theory from this data.

Paper Prototype: A paper representation of a system design, able to simulate operation of that system, which is independent of platform and implementation, and can be used for brainstorming, designing, testing and communication of user interface designs and for identifying usability problems at an early stage of the design process.

Paper-Based Mock-Up: A representation of a specific design idea that is built from simple materials such as paper and cardboard, keeping it cheap and understandable, but making it a physical representation of a design idea for a final system, good for envisioning future products in the very early stages of the design process.

Rapid Ethnography: A collection of field methods to provide designers with a reasonable understanding of users and their activities given a limited amount of time spent in the field gathering data.

Chapter II
Use of Experimental Ethno-Methods to Evaluate the User Experience with Mobile Interactive Multimedia Systems

Anxo Cereijo Roibás
University of Brighton, UK

Stephen Johnson
BT Mobility Research Centre, UK

ABSTRACT

This chapter discusses research initially supported by the Vodafone Group Foundation and the British Royal Academic of Engineering, and subsequently by the BT Mobility Research Centre. It aims to unfold the user experience in future scenarios of mobile interactive multimedia systems, such as mobile iTV with plausible significance in entertainment, work, and government environments. Consolidated and experimental ethnographic data gathering techniques have been used to understand how peripatetic and nomadic users such as commuters and travelers interact in real contexts, taking into account their physical and social environment together with their emotions and feelings during interaction with the system. This approach potentially enhances the consistency and relevance of the results. This chapter also envisages how mobile users could become a sort of 'DIY producers' of digital content, prompting the emergence of mobile communities that collaborate to create their own 'movies' and exchanging them not only with other users but also places (real and virtual environments) and objects (intelligent objects and other digital-physical hybrids). This work illustrates that mobile and pervasive TV would go further than merely broadcasting TV content on handhelds; it will be a platform that will support collaboration and enhancement of creative skills among users.

INTRODUCTION

Interactive TV demands active participation by viewers, and as a result, it considerably affects people's experience with television and their TV-related social behavior. Users' adoption of powerful handhelds with multimedia features, together with an increasing interoperability between platforms, results in the expansion of the iTV consumption beyond the domestic context. We can define this 'almost everywhere TV' as 'pervasive TV.'

The presented research explores realistic and relevant future scenarios for pervasive iTV and for pervasive interactive multimedia systems that address the demands, needs, and desires of a specific category of users: commuters. Likewise, novel processes and structures for content creation, sharing, and consumption that match the nomadic lifestyles of commuters, and embody their values, are investigated.

This research shows that there is, in fact, a growing interest by users in mobile interactive multimedia systems. However, these systems are different from the conventional concept of TV broadcasting on mobile phones. The scenarios that arise in this research are more related to non-professional users co-producing and sharing media content in applications for mobile devices, the internet and iTV for small network communities. According to a recent research by Deloitte Touche Tohmatsu, "Companies have invested significant sums in developing mobile television services so far, but mobile television has had muted commercial impact. Its disappointing performance is likely to continue in 2007. A key reason for this will be weak consumer demand" (The 2007 edition of DTT TMT Industry Group's Telecommunications Predictions, 2007).

There are several research projects addressing different aspects of pervasive interactive multimedia systems and distributed multimedia systems and services (Arreymbi, 2006; Butscher, 2006). Many of them explore either the area of mobile and pervasive games (Barrenho 2005; Capra et al., 2005) or the experimental interactive arts (Frisk 2005).

Furthermore, much of the current research focuses on specific interactive aspects such as the screen (Pham, 2000; Zheng, 2005), the sound (Scheible, 2005), or the digital content in general (Goularte, 2004), but disregards the influence of the context. However, the scope of this project is to analyze the user experience (UX) in a holistic way in order to understand which elements and applications of interactive multimedia systems are suitable in specific contexts, providing the user with a high quality experience.

BACKGROUND

Industry has often failed to understand and forecast users' needs and expectations in sectors that are normally characterized by innovation-driven approaches (such as telecommunications and iTV). Many companies developed applications for handhelds or iTV using inappropriate ICT resources that require massive modifications in users' habits resulting in perceptive or cognitive overload. Consequently, the market's response to investments in developing new products (e.g., mobile TV broadcasting) has not been positive to date. Rapid changes in users' habits and technological advances have generated enormous uncertainties and call for innovative research and development methodologies. As the aspects that need to be considered here have a diverse nature, a cross-disciplinary approach that includes human factor studies, behavioral theories, socio-cultural and economic trends, technological developments and emerging technologies markets, interactive arts, product design, and so forth is necessary. Moreover, several techniques such as collaborative and user-centered approaches that focus on users' cultural, social, behavioral and ergonomic backgrounds must be combined.

Many network operators in Europe, the USA, Japan, Korea and Canada are starting to broadcast TV on handhelds (see Table 1)[1]. This is commonly defined as mobile TV.

There are several reasons that might undermine the success of such operations. The first one is related to the intrinsic physical diversity between

Table 1. Commercial and trial mobile TV launches worldwide

Operator	Country	Platform	Channels	Trial/ Commercial
Bell Mobility	Canada	MobiTV	8	Commercial
Rogers	Canada	MobiTV	9	Commercial
TELUS Mobility	Canada	MobiTV	7	Commercial
Sonera & Elisa	Finland	DVB-H	9	Trial
SK Telecom	South Korea	S-DMB	9	Commercial
O2	UK	DVB-H	16	Trial
Orange	UK	MobiTV	9	Commercial
Virgin Mobile	UK	DAB	3	Trial
Cingular	US	MobiTV	23	Commercial
Midwest Wireless	US	MobiTV	23	Commercial
Sprint	US	MobiTV	23	Commercial

both interfaces (TV and handhelds) making them unsuitable for the same way of delivering of content. The second regards the context of use: TV is traditionally used in a domestic private environment (Spigel, 1992) and usually involves social sharing (Morley, 1986), while mobile phones are mainly used in public environments and entail an individual experience (Perry et al., 2001). Moreover, users are becoming more and more nomadic spending less time at home and in the office. This implies an increasing need for performing our daily tasks while on the move (Leed, 1991). Therefore, unlike TV, handhelds are regularly used in different situations and with different purposes (they are likely to be used as an auxiliary tool to assist users' in a main activity (Harper, 2003). In addition mobile services can be related to the specific context of the user (context awareness). Finally, there are operability differences: TV (including interactive TV) is considered a passive or low interactive medium while handhelds typically demand high interactivity and connectivity. These dissimilarities influence the way with which users interact with the medium and therefore necessitate distinct interaction patterns and content as well as different service formats and features.

MAIN FOCUS OF THE CHAPTER

One of the main assumptions of this work is that commuters are a particularly relevant and interesting population segment for investigating novel processes and forms of interaction with mobile multimedia content. The growing interest and dedication to mobility and mobile life among commuters is strikingly apparent through the rapidly increasing share of resources used for this lifestyle. Use of public transportation, bicycles, and walking in urban mobility[2] has been increasing steadily, as has the amount of money spent for mobility and telecommunications (Pooley et al., 2005). These target users have been further divided into two categories (with a balanced representation of different cultural backgrounds and professional areas and roles): 18-35 year olds and 35-60 year olds.

Furthermore, this work has been influenced and guided by the following premises: small mobile devices can provide both a functional and an effective interactive experience, being able to recreate an enjoyable immersive environment for the user; they are also appropriate for the creation of (and interaction with) new forms of multimedia content and finally, they are suitable tools for context awareness applications.

METHODOLOGY

This work integrates a variety of approaches to evaluate and understand the user experience. These methods include time studies of user panels, observation, mapping of movements and other ethnographic techniques in order to answer the factual questions about the UX in future scenarios of pervasive iTV, interpret the meaning of the findings and describe the relations between more levels of empirical experience and analytical outcome.

In-situ evaluation techniques have been used in several projects to assess the design of interactive systems in public or semi-public environments such as the evaluation of ambient displays at work and in a university (Mankoff et al., 2003); the evaluation of ambient displays for the deaf that visualize peripheral sounds in an office environment (Ho-Ching et al., 2003); the evaluation of a sound system to provide awareness to office staff about events taking place at their desks (Mynatt et al., 1998) and the evaluation of a system of interactive office door displays that had the function of electronic post-it notes to leave messages to the office occupant when they are not there (Cheverst et al., 2003).

Simulations and enactments are very useful when the usage contexts make the mediated data collection particularly difficult due to high privacy, technical, or legal issues (e.g., military environments) or when the system is at a very experimental level. Simulations using proof-of-concept mock-ups or explorative prototypes in labs have been largely used to evaluate the usability and accessibility of interactive systems. Although they might provide valuable information about the UX with a certain interface, they tend to disregard the contextual and emotional aspect of the interaction. Additionally, they can only be used when the conceptual model of the system reaches an adequate level of maturity as they presume the use of a functional prototype.

Such research needs to combine experience, data, analysis, and evaluations from many perspectives in order to achieve a multi-disciplinarily built platform for understanding how and why specific concrete needs, the demand for specific services and technological and aesthetic solutions are integrated in users' social, cultural and aesthetic practices; in short how these shifting trends among commuters evolve and shape. The work has been divided into three main phases.

The first phase is devoted to the analysis of the UX in future scenarios of mobile and ubiquitous i-TV and the elaboration of the usage scenarios consisting in the creation of the scripts for the storyboards. The second and the third phase involved enactments and simulations instead of mediated data collection (as in the previous stage). Creation of usage scenarios is a diffused ethnographic technique used to identify requirements and concept assessment, often combined with laboratory evaluation (Carroll, 2000). There are different typologies of scenarios each one being appropriate for a specific scope: activity scenarios (e.g., based on experiential narratives) are useful during preparatory fieldwork early in the design process (created in the first phase and validated in the second one); mock-up scenarios aim to understand how the designed system suits users' activities (used in the third phase); prototype evaluation scenarios aim to evaluate the interface models of the system; integration scenarios simulate the effect of the finished design. The last two categories have not been used in this project, as they require a rather sophisticated working prototype of the system (which was out of the scope of this work).

The first phase consisted of two initial focus group sessions with each of the target groups identified earlier. Each workshop involved 12 participants and aimed to get the user's view about trends on multimedia mobile applications, TV at home and on the move, new forms of content for mobile TV, advanced interaction possibilities and finally, potential interconnections between handhelds and other devices. This activity has been combined with a theoretical investigation of existing technologies together with successful interactive user experiences in other areas (e.g., games, HCI in Space, etc.). This phase also included ethnographic research using cultural probes, questionnaires and naturalistic observation (photo and video recording in-the-field and data analysis). While focus groups and analysis of study-cases were good sources of functional and data requirements; cultural probes and questionnaires provided good information about users' requirements and finally in-the-field observation has been a very valuable technique to identify environmental and usability requirements. Furthermore, the information collected here provided the basis for the scenario scripts that were evaluated in the following stage.

The second phase aimed to validate some significant usage scenarios and subsequently to identify and classify innovative related applications exploring, at the same time, radically new forms of 'smart' and 'malleable' content. This process consisted of two workshops that used role-playing as the basis for a collaborative design approach. It involved twelve representatives of both target groups of users to represent and discuss the scenario scripts (that were elaborated in the previous phase) in order to confirm the legitimacy of the scenarios and experience models proposed in terms of relevance, effectiveness and soundness.

The third phase involved the creation of proof of concept mock-ups and development of user experiments in order to bring to light the feasibility and usability of the scenarios, applications and forms of content previously identified. In this phase some experimental low-fi prototypes of applications were developed (and empirically evaluated in the field) that operate across an integrated system of interfaces and form factors that connote pervasive iTV (typically mobile phones, PCs and iTV). Thirty users aged between 18 and 60, with a peripatetic lifestyle and mixed cultural and professional backgrounds, took part in this evaluation through two sessions in a public plaza.

Focus groups in the first phase of the project provided a framework for discussion about the future use of multimedia content in handhelds in contexts of pervasive communication. During the focus group sessions, the facilitator stimulated brainstorming around the following topics: trends for nomadic users for work, leisure and government ('time spent at home/work/on the move,' 'what to do on the move and with whom'); TV versus mobiles ('What do we enjoy in TV when and why, what we don't enjoy and why,' 'TV on mobile'); novel smart multimedia content ('beyond images, video and sound,' 'poor content vs. rich and smart content,' 'contextualized content'); advanced interaction ('content malleability, gestures, haptics, multi-sensorial devices, holograms, voice navigation'); mobiles in connection with other devices, things and places ('mobile and TV, mobile and the office, mobile and the house, mobile and the street—buildings, objects, events, people'). Participants also provided short 'stories' (experiential narratives) regarding their view on the topic. Some stories were real, some other were fiction.

Cultural Probes (Figure 1) aimed to get inspirational responses to understand beliefs, desires, aesthetic preferences and cultural concerns of users without observing them directly. This technique was initially used by Gaver in industrial design (Gaver et al., 1999) and has recently been exported to HCI (Hulkko et al., 2004). Six selected users were given a cultural probes pack under the condition of completing and returning them after two weeks.

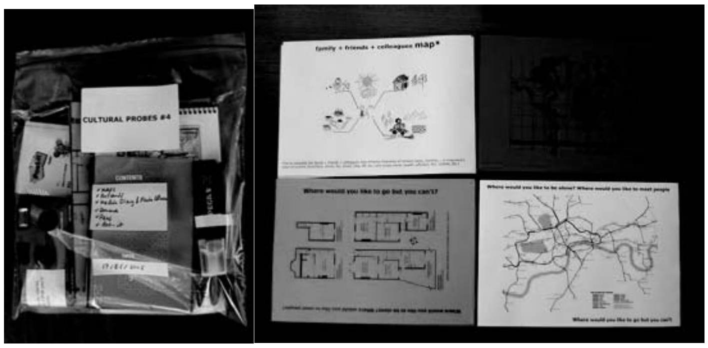
Each pack included four main items with the following instructions:

- **Maps:** World ('where would you imagine having a daydream?'); City ('Where would you like to go now but you can't?'); House ('Where would you like to be alone?', 'where would you like to meet people?'); Family, friends and colleagues relationships ('show frequency and nature of contacts').
- **Camera:** 'take a picture of an image/video you'd like to take with your mobile.'
- **Media Diary:** 'record TV, cinema and radio use (what, when, where, with whom).'
- **Photo album and colour pencils:** 'collect things, images and stories of your week; make sketches.'

Questionnaires (Figure 2) were designed as a set of 11 postcards in order to provide a very informal and open approach, encouraging instinctive and casual replies about the users' vision on the topic and were distributed to twelve target users[3].

Each card has an image on the front, and one question on the back such as: 'which device/s would you take with you if exiled to a desert island?' 'When and where can you feel over-loaded with information?' 'When & where would you like to connect it with other devices, things or places?' 'When and where the use of a mobile phone can be a collective experience (several users sharing the same application using their own devices)? And a public experience (several users operating the same device simultaneously)?' 'How would you feel about having awareness of other devices, people, places and things in your handheld?' 'Name a book or movie with a future scenario you liked and another one you disliked (or tell us about your own idea).'

Figure 1. Cultural probes pack and detail of the maps



User-centered design researchers have historically favored direct observation because it is a method that places the researcher in the context in which technology use occurs (Hagen et al., 2005). However, mobile devices are designed for individual use within a personal body space. Therefore, observing the interface actions of the user can be physically arduous (Kjeldskov et al., 2005; Mark et al., 2003; Newcomb et al., 2003). Naturalistic observation (included note taking, photography and video recording) by following commuters for three months in public places in London such as theatres, parks, stores, stations and squares as well as on public transport between London and Brighton provided a useful insight into the use of mobile phones as multimedia communication tools (mainly taking video and pictures and sending, TV streaming, picture editing, etc.). During this activity, it was also possible to determine and map the behavior of users in common public meeting spaces such as stations and parks. The typical duration over which these maps were determined was in the order of a couple of hours and aimed to be a source of information regarding the sociability processes in public spaces as well as the effect of technology (especially mobile phones) in this phenomenon.

These techniques were useful not only to identify users' main requirements, but also to write scripts for the scenarios. The scenarios were then validated in two sessions, each one with 12 representatives of one of the user categories previously identified. Each session consisted of in-situ theatre performances carried out by the participants themselves following the scripts of the scenarios.

The last part of the project used horizontal proof of concept mock-ups to assess relevant and plausible applications that were identified during the research. This phase consisted of two evaluation sessions, again using role-playing, but this time in the field.

The plays were performed in public environments (mainly squares in the city) by some of the users while the others could comment on what they were watching. The advantage of using in-situ enactments was that they provided precious information about some contextual factors that had not been identified in the research process. Examples of these valuable outcomes included the users' embarrassment of watching a movie in a crowded train and concerns about the phone being stolen when using multimedia applications in public places in the city, and so on.

These very early prototypes incite experimentation, are easy to use and adopt, encourage discussion between users and designers and have a very low cost. However, due to their low-fi appearance they might appear unconvincing, raising criticism by the users. Moreover, they focus excessively on functionality not tested in the real usage contexts. Figure 3 shows one of the experimental prototypes—a mobile 'memory box.'

Figure 2. The questionnaire

Theatre Workshops: Personas and Scenarios (Phase Two)

Theatrical performances are a valuable technique that can be used to collect data and identify requirements and other crucial information from users such as feelings and emotions. This is because the user's self-esteem is not directly under scrutiny and so inhibiting factors like embarrassment or shyness are less likely to be elicited.

The use of drama (Iacucci et al., 2000; Sato, 1999) can be an effective tool in participatory design as it facilitates dialogue between designers and users. According to Newell, it can cross boundaries of technical language and knowledge, allowing elderly potential users to be involved effectively in the process of design at the pre-prototyping stage (Newell et al., 2006).

Live play gives the audience the possibility of interacting directly with the actors, providing feedback about the feasibility and realism of the situation played, but in the case of budget restrictions the use of video can be an effective alternative.

Three personas (Table 1) have been created to represent the three main typologies of users of these pervasive multimedia systems. Paul and Marina embody the two categories of users directly involved in this research. The third one, Alex, can be included in the younger typology, and represents a driver category for the diffusion of these systems: techno-fun users. Further users' profiles could have been considered, however according to the authors' experience in these sorts of projects, the benefit of increasing the number of personas in such systems is uncertain.

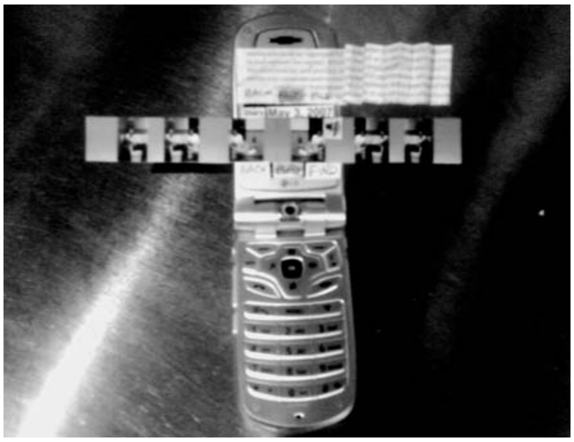
The scripts for the three following scenarios have been created from the stories[4] that participants brought to the focus groups (Howard et al., 2002). Each scenario corresponds to a prospected UX with pervasive multimedia systems for each user profile. The first scenario provides a more conservative approach to the problem, while the third one gives a more futuristic view of the prospective system. Although some elements that have been included in the scenarios are not novel individually, the intention was to recreate a complete view of how pervasive interactive multimedia systems could be totally embedded in the personas' daily lives. The following Summarizes the scripts for each one of the scenarios:

First Scenario: "Paul"

Morning routine: Paul has breakfast with his family whilst watching the morning news on the digital tablecloth. He then drives his son to school after which his wife gets an automatic notification (voice message on her mobile or digital tablecloth) that the boy has arrived at school. While driving to his office, Paul receives personalized local video news on his in-car-navigator.

Work routine: All of Paul's mobile communications go across his desktop when he is sitting in his chair at work. He uses the local satellite system to find a picnic area, the location of which he forwards to his wife in the form of a multimedia (video and map) message inviting her to lunch. After Paul leaves the office to meet her, he gets a video message on his in-car-navigator from a partner asking urgently for an important file. Paul retrieves the file in one of his company's folders using the voice interaction system, and replies to the message. He then sets the work message system to very-urgent-only mode since he has the afternoon off. He then sees by his wife's position on his navigator map that she has arrived at the arranged the meeting point for lunch.

Figure 3. Proof-of-concept mock-up for the 'memory-box'



Social lunch: As Paul joins his wife he receives an info alert message, which he accepts to get a short video clip about a nearby landmark. Using voice activation, Paul uses his mobile to provide him with related information about the area and other subjects of interest. Finding digital messages from a relative that has been to this landmark before, Paul and his wife view some of them and then leave one of their own for others to discover later. Following lunch, he notices in his phone's EDG (electronic diary guide) that his aged uncle is engaged in a global warming activist's discussion through videoconferencing. He decides to join and support him by providing some video documentary evidence of a dry lake nearby.

Afternoon leisure: Paul checks the state of the river through the rowing club live-cam and finds the water is too rough for rowing. He sends a link with the live video of the river to his pals together with a message suggesting they reschedule their plans to row together. Having changed his schedule, he visits a nearby market mentioned in the morning news and finds an interesting but pricey item. He starts a videoconference with a knowledgeable friend who examines the piece via Paul's mobile cam and offers an opinion on the value of the item. Unconvinced he should buy it, Paul uses a matchmaking system to locate similar collectors who are visiting the market and he finds two people. A quick look at their profiles tells him they are worth a short meeting so he sends them a message inviting them for a coffee. Meanwhile, Paul's wife checks her virtual map, spots a nice gallery nearby, and makes a short virtual visit. When she notices on the map that one of her friends is nearby, she sends an invitation for her to visit the gallery.

Evening routine: On the way back home, his wife uses her mobile shared-whiteboard with her best friend to help organize a party by choosing the menu, guests, and the decoration. After picking up their son from school, she receives a notification that her favorite TV show is about to begin so she remotely checks that the home video recorder has been set properly. She then accepts to watch a summary of today's episode on the in-car-navigator, which she also forwards to her friends.

Second Scenario: "Marina"

Morning routine: Waking up in a Shanghai hotel, Marina uses her mobile to activate personalized BBC news displayed on the interactive portrait near her bed. Although Mum is not feeling well these days, Marina notices through her EDG that her Mum is video-sharing a recipe with her cooking online community; this relieves her. A second message pops up to remind Marina of breakfast plans with a friend near the Yuyuan Garden.

On the move: On the bus heading to the café, Marina follows an electronic-paper map of the route in order to familiarize herself with the city. The live interactive local map includes links to helpful information. The map notifies her of digital messages attached to particular landmarks (when passing by), and provides the opportunity for Marina to leave her own comments for others to see. Passing a concert hall, Marina accepts to download a video clip informing her of an upcoming performance by her favorite band. Immediately she books using her mobile and sends the video message to two friends, inviting them to join her. The map shows her that there is one of her jogging partners in a café nearby, but she decides to disregard the notice as she is in a rush to her appointment.

Table 2. Summarized description of the three main personas

Paul is a married 52 years old mechanical engineer who has just moved to the countryside near Edinburgh and has 2 sons: one still in school while the other has just started at the university. Paul enjoys fiction movies, rowing, gardening, and collecting. **Marina**, a single 28 years old lawyer living in east London and frequently travelling abroad for work, is interested in archaeology, travelling and jogging. **Alex**, a 21 years old sociology student living in Cardiff, has a girlfriend in Copenhagen so he travels there very often. He is passionate about music, surfing and clubbing.

Social interaction: Still on the bus, Marina notices an attractive man reading a book on a nearby seat. Pointing her mobile toward the book, it reveals the book's title which she finds fascinating and increases her interest in meeting the person reading it. She approaches the man, introduces herself, and informs him that she is about to get off the bus but would like to talk with him another time. Shaking their mobiles, they exchange their contact information with each other. After leaving the bus, Marina checks the guy's data on her mobile and discovers it includes personal information about his interests, including video galleries, stories, and more, which reveal they have many points in common.

Context interaction: As Marina walks across the Yuyuan garden, a work of art grabs her attention. A quick query using her mobile (thanks to a pattern recognition search engine system) informs her of the artwork's significance and of a projected interactive discussion board related to it. Marina chooses to leave her own digital 'waymark' and to create a digital video card for her friends. Realizing she is now late for breakfast, Marina notices in the interactive map that her friend is already there, so she hurries.

Social interaction: While sitting in the café, Marina and her friend decide to go on holiday together. Using the electronic map they check travelers' advice and recommendations for prospective locations which includes travelers' self-authored audio-visual content about their experiences. Joining their two mobiles they double their screen size and access enhanced navigation features.

Third Scenario: "Alex"

Morning routine: Alex wakes up in his girlfriend's apartment in Copenhagen to find Linda has already left for work. Pointing his finger toward the electronic wallpaper, he voice-activates a holographic 3D videoconference with Linda. Using his mobile he activates a holographic projection of MTV-TRL to view whilst he starts his day. As he would like to go surfing, Alex requests activation of live-cams of other surfers' handhelds so he can check weather conditions. The only surfer on the beach accepts to tele-activate his cam, revealing a calm sea unsuitable for surfing, so Alex cancels his plans and heads out for breakfast. In a quiet café he catches up with news from his favorite mobloggers and participates in a video debate about a recent sailing race. He receives a live-cam activation request from a viewer of a reality TV show to which Alex is subscribed. Viewers of the show can access his mobile or home cam when they want.

Digital geo-caching: An alert on his mobile prompts Alex to receive a video message that turns out to have mysterious and cryptic significance.

With some research, Alex solves the riddle and gets the prize of a virtual ark on his mobile. An experienced player of this game, Alex knows he now has 24 hours to find a new hiding place for the ark and create a new riddle for the next person to solve. He hops on a bus heading across the Øresund Bridge to Malmö, Sweden, and streams a short video of the sky, sea, and coastline to the other players. While he searches for a good location to hide the ark, he uses his mobile to deposit multimedia messages on different landmarks on his way giving hints for other players to find the ark.

Romantic gesture: Back in Copenhagen and thinking of Linda stuck in a meeting at work, Alex records a video of himself with a romantic quote. He then buys a bunch of daisies and digitally attaches the video to the bouquet which he leaves with the receptionist at Linda's office. Later, when Linda receives the flowers, the registered video message appears on her digital book.

Virtual memory: Alex senses he is forgetting an obligation. He checks his electronic agenda but nothing is scheduled. When he rewinds his personal memory box system he recalls the previous night he had promised to make a video of a local band and narrowcast it to the interested people of the MTV TRL network. He adds this to his electronic agenda. Then Alex uses his mobile to check how many clips from his personal surf video gallery have been sold recently. Disappointed in what he finds, he records a note to himself to do more interactive advertising.

These scenarios have subsequently been assessed in two workshops through dramatizations. Some of the workshop participants were asked to act out the scripts to represent the scenarios. In this way, users and designers were able to discuss the feasibility of the different elements of the system proposed. This methodology has proved to be useful not only to confirm their reliability and relevance but also, to achieve a high engagement of the users during the design process (Newell et al., 2006). Some elements of the scenarios have not been intentionally described in detail (e.g., the EDG), because it wasn't enough information in the previous phase to understand how the users have envisaged these applications. Only if during the scenario dramatizations the participants found these applications relevant, designers generated discussion with the users in order to define them better.

During the role-play in the assessment workshops some applications presented here (e.g., the digital table cloth in the first scenario) were strongly criticized by the users and therefore they have not been considered in the development of the proof of concept mock-ups. The EDG produced general excitement among the participants, as they saw it as an effective tool for obtaining information about the wellbeing of older relatives and friends without letting them feel observed, thus supporting independent living and encouraging mobility. Other examples of applications described in the scenarios have been amended or refined in order to better satisfy the users' requirements (e.g., the interactive map and the in-car communication system).

Experimental Prototypes (Phase Three)

The scenarios provided the basis for identification of possible implementation settings and verification of system requirements. Once potential interfaces and applications were determined, assessing their quality in-the-field required the development of proof of concept mock-ups. Since this work focused on the nomadic and peripatetic behavior of users, ambient and home interfaces have not been considered. A description of the different applications that have been assessed in the form of low-fi experimental prototypes are detailed in the following paragraphs. Many of the devices and applications proposed are not original, however collectively they combine to provide a highly innovative pervasive interactive multimedia system. In fact, if any of the identified interfaces or applications had already been developed (e.g., a multimedia mobile phone), we would have considered it appropriate to incorporate it in the system instead of creating a new one. The proof of concept mock-ups was made by combining paper, cardboard, and real mobile-phones. The envisaged pervasive interactive multimedia system comprised of five interfaces; a mobile phone, an interactive map, public interactive displays, an in-car multimedia system and a memory pin.

Handheld

This mobile device has a traditional clamshell design with a pivotable color display, photo and video camera, and keypad-based standard interaction as well as voice-based interaction.

In addition, a small transmitter inside the device enables it to serve as a pointer and allows interaction with TV screens, public digital displays (much like how a mouse is used to point to a computer display) and with intelligent objects such as bus stops (to get information from them).

Also envisaged is the possibility combining two devices to double the size screen and permit enhanced navigation, which would increase the sociability potential of this application.

Applications for this device include:

- Context aware infotainment such as local video news, visualization of user's position on a map, reception of in-situ multimedia alert-messages from things, places (landmarks, building, etc.) and events.
- Distance vision such as remote cam activation and control (zoom, positioning). The remote can be an autonomous device or can be embedded in other users' handhelds.
- Customized multimedia content such as a embedded mobile live-encyclopedia with a pattern recognition search engine (linked to TV-video content).
- Self-authoring system enabling recording, editing and sharing, broadcasting or narrowcasting of personal videos, life TV video debate in videoconferencing, mob-blogging and co-production of reality-TV channels, notification of a live event or TV series and video clip summary of a TV show and possibility of storing, editing or sharing it to other users.
- 'Memory box' that enables users to register self-authored multimedia content (memories) that would be delivered in specified future occasions such as birthdays or graduation-day to their younger relatives or friends, to keep their presence alive even when they have passed away.
- Socialization and social awareness system such as a matchmaking system. It can be used to find users with desirable profiles, shared whiteboards or locator of buddies. New interaction models for this application include the possibility of pointing the mobile towards a person to get info about them, and exchange of personal information by shaking hands between users.
- Electronic diary guide where users can manage their daily appointments as well as checking whether their friends or relatives are currently engaged in any sort of interaction with the system (e.g., discussion group, etc.).

Public Interactive Display

Interactive digital billboard that displays customized info based on profiles of passers-by with a pointing-based interaction system and voice-based interaction capabilities. Although similar interfaces have been already developed in other projects, the scenarios showed that it was crucial in terms of completeness to integrate it in the proposed system (McCarthy et al., 200; Russell et al., 2002).

Applications for this device include:

- Storage and display of digital messages from people that have been there before, public interactive discussion board (using one's mobile as both pointing device and content editing tool), display of incoming personal messages (video, text or voice through sonic cones for direct sound to the user).

Personal Interactive Map

This device is best described as electronic foldable paper with full touch-screen display, voice based interaction and GPS location based system (see Figure 4).

Applications for this device include:

- Mapping and routing services such as local maps and interactive ads
- Social and context awareness such as location of the user, other people, things and places, routing system with multimedia information about nearby people, things and places including an instant sessaging system (IM).
- Micro-payment e-commerce applications such as the possibility of making bookings or purchases related to the above (e.g., concert) and forwarding the info to someone else.
- Display of self-authored content such as users edited travel guides and maps.

In-Car Multimedia Communication System

When the user sits in the driver's seat of a car with this built-in system, all of their mobile communications automatically route through it. The system consists of an adjustable monitor and semitransparent projection (for the driver's use) in the front glass (Figure 5), touch-screen capabilities, and voice-based interaction.

Applications for this device include:

- Context aware infotainment such as local video news, local satellite maps (and possibility of forwarding the location to someone with a voice or text message).
- Social awareness such as location of other people.
- *Busitainment* applications such as video message and videoconference system, retrieval of personal or work files, automatic addressing of messages, different 'screening' modes: family, personal, work and very-urgent-only.

Memory Pin

This device is simply a small, low-cost storage container capable of interacting with a user's mobile on request in order to store a wireless download of self-authored multimedia content. The 'pin' is then attached to a desired object (as shown in Figure 6) and ready to upload its multimedia content to nearby devices (PID, e-paper).

Applications for this device include:

- Download of text, sounds and movies from a mobile.
- Upload of the registered video-quote on a user's device (digital book, PC, PID, electronic paper) by simple touch.

The low-fi prototypes representing these interfaces and applications have been tested in-the-field. However, to overcome safety concerns in the case of the in-car system, the assessment was performed in a motionless vehicle. This unrealistic testing context has made questionable the validity of the results. However, the semitransparent GUI projection on to the front windscreen did elicit users' worries about safety and security.

In the other cases, when the evaluation considered the real context of use, they provided crucial information about how the physical and social environment can influence the use of the system. For example, the interactive map raised concerns about the management of the privacy, and the mobile multimedia phone about embarrassment in crowded areas. During the in-the-field assessment of the proof of concept mock-ups, the experience with some applications (like the 'memory box') was highly praised.

RESULTS ANALYSIS

In order to address complex issues such as understanding, emotion, security, trust and privacy, the data gathering techniques presented in this chapter focus on users rather than on their tasks or objectives with the analyzed interfaces. This research shows how the physical and social contexts have a strong impact on the users' attitudes towards mobile interactive multimedia applications: the context influences the users' emotions and feelings towards the interaction process, persuading or discouraging its use (Kjelskov et al., 2004). For example, during the in-the-field assessment of the proof of concept mock-ups some users felt unsafe recording video with their mobile phones in a crowded street, as they were very concerned about theft.

The questionnaires, observations and the focus groups revealed two main users' categories when considering the creation and sharing of self-authored multimedia content:

- Spontaneous or impulsive user (e.g., when travelling, during an exciting night out, when observing an interesting thing, place, or performance or just to update about domestic issues such as children, new partner, etc.). The addressees are the members of the user's restricted social personal circle: family, friends and colleagues.
- Reiterative or structured user (e.g., mob-blogs). The addressees belong to a broader social circle such as enlarged communities.

The cultural probes showed a clear desire by users for using their handhelds to create self-authored video content for two main purposes: as an enhanced democratic tool (e.g., voting on public issues or having 'five minutes of glory in

Figure 4. Interactive map from the Vodafone Futures

TV') and to leave their 'signature' along their way (e.g., by putting down personal-digital content on public digital board at monuments or other places). Applications regarding exchange of multimedia content with objects and places have been explored in many different contexts: visiting a city (Brown et al., 2005; Cheverst et al., 2003), playing pervasive games (Benford et al., 2004), leaving signs and building communities (Burrell & Gay, 2002; Giles & Thelwall, 2005; Persson et al., 2002).

The probes also exposed users' preferences when receiving multimedia content on their handset from people, places or things: 'If on the move, it's better if related to my context.' Context awareness provides customized information that can be defined as the right information in the right place and in the right time. In this sense it is interesting to note the work of Abowd and Mynatt (2000) who apply a set of five questions to obtain what they call a good minimal set of necessary context.

Observations and mapping of movements provided qualitative information about how the social context influences the use of mobile phones in public spaces. For example, it revealed how mobile phones encouraged their owners to temporarily disconnect from a social group (such as a group of friends) during the period of use (for example, whilst reading a message).

The experiential narratives that participants presented during the focus groups, raised the following issues for mobile interactive multimedia systems: sociability (e.g., to allow users traveling together to share the experience of viewing a video with their mobile phones) and collaboration (enable users who are in different places to exchange moods, share information and even work together) (Lull, 1980), context awareness (both services and content should be customized and related to the specific users' context), creativity (enable nomadic and peripatetic users to produce self-authored multimedia content), interactivity (interfaces need to support a high level of interactivity, by using new modalities such as gestures) (Palen et al., 2000), convergence (enable users to use the most plausible and appropriate interface in each context: iTV, mobile phones, in-car-navigators and

Figure 5. Semitransparent projection in the car windshield

the Internet) and connectivity (enable different ways of communication among users: one-to-one and one-to-many).

Surprisingly, this research highlighted very little appeal by users in receiving broadcasts of traditional TV formats on their mobile phones (except some exceptions such as brief live updates of a decisive football match or extraordinary news). Therefore the concept of mobile or pervasive iTV is more likely to be related to the emergence of mobile communities that support 'DIY producers' of multimedia content: they will create multimedia content in specific contexts, with precise purposes and share it with others.

The interfaces and applications show how such open, diffuse, and pervasive interactive multimedia systems provide an exceptional virtual platform that might foster and enhance the development of new communities of creative users that can share moods, content and collaborate with different purposes such as work, entertainment or government. In the specific case of entertainment it is worth mentioning Davenport's view of the topic: 'Since the earliest days of cinema, artists and technologists have dreamt of a future in which everyone could create and share their vision of the world. With the evolution of ubiquitous mobile networks and the enhanced mobile handset as creative device, we are on the cusp of realizing improvisational media fabrics as an active expression in our daily lives'.[5]

CONCLUSION

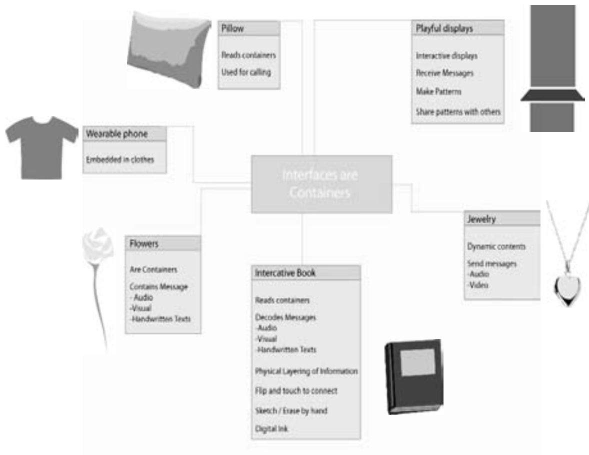
Traditional data gathering and evaluation techniques based on cognitive psychology focus on the human machine interaction and disregard a crucial aspect in the process: the context of the user. The physical and social context might have a strong impact in the use of the analyzed interfaces: it influences in a positive or negative way the users' emotions and feelings towards the interaction process, persuading or discouraging its use.

This research tries to recognize the mutual influence between technology and society. Just as technology shapes society, we also need to investigate how society shapes technology. This particularly holds true with a social technology that needs to be integrated with household routines. In making predictions about new technology we need to explore the critical disconnections between the ways in which such technologies are produced and the ways in which they are consumed, naturalized and rejected (Fischer, 1992; Lee & Lee, 1995).

Handsets are becoming tools for creation, editing and diffusion of personalized and personal multimedia content and this attribute allow users to become 'DIY producers' of digital content (Cereijo Roibas & Sala, 2004). Users will be able to create their own multimedia content to share with others. Therefore new communities of nomad and peripatetic users will find themselves in original communication contexts and in novel expressive situations: they will be able to create their own 'movies' and share them with other users, places (real and virtual environments) and objects (intelligent objects and other digital-physical hybrids). This expression of users' creativity in pervasive interactive multimedia systems needs to be corroborated by interfaces that support some form of users' creativity, collaboration social interaction (Ducheneaut & Moore, 2004; Nardi et al., 2004; Preece & Maloney-Krichmar, 2003).

The research methodology adopted helped to confirm the validity of the three premises that guided this work. As shown in the second and third scenarios, small mobile devices can provide both a functional and an effective interactive experience. As it has been recently corroborated by the

Figure 6. Memory pin: Interfaces as containers of information



success of the Wii, the fifth video game console released by Nintendo (European consumers snap up 325,000 Wii consoles in two days. 2006), they are also able to recreate an enjoyable immersive environment for the user. The distinctive feature of the new game console is its small wireless controller, the Wii Remote, which can be used as a handheld pointing device and can detect motion and rotation in three dimensions providing the gamer with a high immersive experience.

While the first and second scenarios make it particularly clear how handhelds are appropriate for the creation of new forms of multimedia content by the user, the third, however, provides evidence of how novel forms of interaction could be implemented (see Romantic Gesture). The recently unveiled Apple iPhone promises deliver a fine example of novel interaction with content through its multi-touch screen (that uses users' fingers as the ultimate pointing device) and multi-touch sensing system. One of its innovative interactive features is the ability to zoom objects in and out such as photos, by placing two fingers on the screen and moving them farther apart or closer together as if stretching or squeezing the image (Apple's 'magical' iPhone unveiled. 2007).

All of the three scenarios showed how mobile phones are suitable tools for context related content and context awareness applications. The enhanced social interaction shown in the second scenario is particularly interesting.

Our future work will look at the exploration of how new technological paradigms will affect the perceived quality of experience in pervasive interactive multimedia systems. These paradigms include hybrid artifacts, use of biotechnology, advanced interaction modalities, new forms of content and novel intelligent environments, immersive environments such as collaborative virtual environments and multi-user environments. In this sense, an examination of the contributions that disciplines such as the interactive arts, space technology, medicine and games could give to this area might offer significant insights. At the same time, to achieve an adequate outcome in these new territories, we will also look at new immersive field research instruments that reduce, as much as possible, bias and subjectivity in ethnographic research (LeCompte, 1987).

REFERENCES

Abowd, G. D., & Mynatt, E. D. (2000). Charting past, present and future research in ubiquitous computing. ACM transactions on computer-human interaction, *ACM Press*, *7*(1), 29-58.

Apple's 'magical' iPhone unveiled. Retrieved on January 9, 2007 from http://news.bbc.co.uk/1/hi/technology/6246063.stm

Arreymbi J., Gachanga, E. (2006). Interactive design and delivery challenges for wireless handheld multimedia systems. *International Conference on Internet Computing* (pp. 377-386).

Barrenho, F., Romão, T., Martins, T., & Correia, N. In Authoring environment: Interfaces for creating spatial stories and gaming activities. In *Proceedings of ACM SIGCHI International Conference on Advances in Computer Entertainment Technology (ACM ACE 2006).* New York: ACM Press.

Benford, S., Flintham, M., Drozd, A., Anastasi, R., Rowland, D., Tandavanitj, N. et al. (2004). Uncle Roy all around you: Implicating the city in a location-based performance. In *Proceedings Of ACM Advanced Computer Entertainment (ACE 2004*). Singapore: ACM Press

Brown B., Chalmers M., Bell M., MacColl I., Hall M., MacColl I. et al. (2005) Sharing the square: collaborative leisure in the city streets. I*n Proceedings of ECSCW 2005* (pp. 427-429). Paris: Springer

Burrell J., & Gay, G.K. (2002). E-graffiti: evaluating realworld use of a context-aware system. *Interacting with Computers, 14.*

Butscher B., Moeller E., & Pusch H. (1996, March 4-6). *Interactive Distributed Multimedia Systems and Services.* European Workshop IDMS '96. Berlin, Germany.

Capra M., Radenkovic, M., & Benford, S. (2005). Multimedia challenges raised by pervasive games. In *Proceedings in ACM Multimedia 2005.* New York: ACM Press.

Carroll, J. M. (2000). *Making use of scenario-based design of human computer interactions.* Cambridge, MA: MIT Press.

Cereijo Roibas, A., Sala, R. (2004). Main HCI issues for the design of interfaces for ubiquitous interactive multimedia broadcast. *Interactions Magazine,* 51-53.

Cheverst, K., Dix, A., Fitton, D., & Rouncefield, M. (2003). 'Out to lunch': Exploring the sharing of personal context through office door displays. In *Proceedings of the 2003 Australasian Computer-Human Conference, OzCHI 2003* (pp. 74-83), S. Viller and P. Wyeth (Eds.), Canberra: CHISIG.

Ducheneaut, N., & Moore, R. J. (2004). The social side of gaming: A study of interaction patterns in a massively multiplayer online game. In *Proceedings of CSCW 2004* (pp. 360-369). Chicago, IL: ACM Press.

European consumers snap up 325,000 Wii consoles in two days. *GamesIndustry.biz.* Retrieved December 13, 2006 from http://www.gamesindustry.biz/content_page.php?aid=21691

Fischer, C. S. (1992). *America calling: A social history of the telephone*. Berkeley, CA: University of California Press.

Frisk, H., & Yoshida, M. (2005). New communications technology in the context of interactive sound art: An empirical analysis. *Organised Sound*, *10*(2), 121-127.

Gaver, W.W., Dunne, A., & Pacenti, E. (1999). Cultural probes. *Interactions, 6*(1), 21–29.

Giles L., & Thelwall, S. (2005). *Urban tapestries: public authoring, place and mobility.* London: Proboscis.

Goularte, R., Cattelan, R. G., Camacho-Guerrero, J. A., Valter, J., Inacio, R., & Pimentel, M. G. C. (2004). Interactive multimedia annotations: enriching and extending content. In *ACM DocEng '04* (pp. 84–86).

Hagen, P., Robertson, T., Kan, M., & Sadler, K. (2005). Emerging research methods for understanding mobile technology use. In *proceedings of OZCHI 2005* (pp. 1-10). Canberra, Australia.

Harper, R. (2003). People versus information: The evolution of mobile technology. In L. Chittaro (Ed.), *Human computer interaction with mobile devices* (pp. 1-15). Berlin: Springer.

Ho-Ching, F.W., Mankoff, J., & Landay, J.A. (2003). Can you see what I hear?: The design and evaluation of a peripheral sound display for the deaf. In *Proceedings of ACM CHI 2003* (pp. 161-168).

Howard, S., Carroll, J., Murphy, J, & Peck, J. (2002). Using endowed props in scenario-based design. In *NordiCHI 2002*.

Hulkko, S., Mattelmäki, T., Virtanen, K., & Keinonen, T. (2004). Mobile probes. In *proceedings of the 3rd Nordic Conference on Human-Computer Interaction* (pp. 43-51). Tampere, Finland: ACM Press.

Hutchinson, S. (2001). Urban moving 2030: transport typologies for the future city. In J. Myerson (Eds.), *The Helen Hamlyn research associates catalogue 2001* (p. 22). London: Helen Hamlyn Research Centre.

Iacucci, G., Kuutti, K., & Ranta, M. (2000). On the move with a magic thing: Role playing in concept design of mobile services and devices. At DIS 2000.

Kjeldskov, J., & Stage, J. (2004). New techniques for usability evaluation of mobile systems. *International Journal of Human-Computer Studies, 60*, 599-620

Kjeldskov, J., Graham, C., Pedell, S., Vetere, F., Howard, S., Balbo, S. et al. (2005). Usability of a mobile guide: The influence of location, participants and resource. *Behaviour and Information Technology, 24*, 51-65.

LeCompte, M.D. (1987). Bias in the biography: Bias and subjectivity in ethnographic research. *Anthropology and Education Quarterly, 18*(2), 43-52.

Lee, B., &Lee, R.S. (1995). How and why people watch tv: Implications for the future of interactive television. *Journal of Advertising Research, 35*(6).

Leed, E.J. (1991). *The mind of the traveller.* New York: Basic Book.

Lull, J. (1980). The social uses of television. *Human Communication Research, 6*(3).

Mankoff, J., Dey, A. K., Hsieh G., Kientz, J., Lederer, S., & Ames, M. (2003). Heuristic evaluation of ambient displays. In *Proceedings of CHI 2003*.

Mark, G., Christensen, U., & Shafae, M.(2001). A methodology using a microcamera for studying mobile IT usage and person mobility. In *Proceedings of CHI 2001.* Seattle, WA.

McCarthy, J.F.,Costa, T.J., & Liongosari, E.S. (2001). UniCast, OutCast & GroupCast: Three steps toward ubiquitous peripheral displays. *UBICOMP 2001*, Atlanta.

Morley, D. (1986). *Family television. Cultural power and domestic leisure*. London: Comedia.

Mynatt, E.D., Back, M., Want, R., Baer, M., & Ellis, J. B. (1998). Designing audio aura. In *Proceedings of CHI '98 (Los Angeles, California)* (pp. 566-573). ACM.

Nardi, B., Schiano, D., & Gumbrecht, M. (2004). Blogging as social activity, or, would you let 900 million people read your diary? In *Proceedings of CSCW 2004* (pp. 222-231). Chicago, IL: ACM Press.

Newcomb, E., Pashley, T., & Stasko, J. (2003). Mobile computing in the retail arena. In *proceedings CHI 2003.* Ft. Lauderdale, FL: ACM Press.

Newell, A. F., Gregor, P., & Alm, N. (2006). Theatre as an intermediary between users and CHI designers. *CHI 2006*. Montreal, Quebec, Canada.

Palen, L., Salzman, M., & Youngs, E. (2000). Going wireless: Behavior of practices of new mobile phone users. In *Proceedings CSCW 2000* (pp. 201-210).

Perry, M., O'Hara, K., Sellen, A., Harper, R., & Brown, B.A.T. (2001). Dealing with mobility: understanding access anytime, anywhere. *ACM Transactions on Computer-Human Interaction (ToCHI), 4*(8), 1-25.

Persson, P., Fagerberg, P. Geonotes: *A real-use study of a public location-aware community system* (Tech. Rep. T2002:27, ISSN 1100-3154).

Pham, T-L., Schneider, G., & Goose, S. (2000). *A situated computing framework for mobile and ubiquitous multimedia access using small screen and composite devices.* New York: ACM Press.

Pooley, C. G., Turnbull, J., & Adams, M. (2005). *A mobile century? Changes in everyday mobility in Britain in the twentieth century*. Ashgate, Aldershot, Hampshire.

Preece, J., Maloney-Krichmar, D. (2003). Online communities: Focusing on sociability and usability. In J.A. Jacko & A. Sears (Eds.), *2003: Handbook of human-computer interaction*. London: Lawrence Erlbaum Associates Inc.

Russell, D. M., Drews, C., Sue, A. (2002). Social aspects of using large public interactive displays for collaboration. Ubicomp 2002.

Sato, & Salvador. (1999). Playacting and focus troupes: Theatre techniques for creating quick, intense, immersive, and engaging focus groups. *Interactions, sept/oct.*

Scheible J., & Ojala, T. MobiLenin—Combining a multi-track music video, personal mobile phones and a public display into multi-user interactive entertainment. In *Proceedings ACM Multimedia 2005* (pp. 199-208). New York: ACM Press.

Spigel, L. (1992). *Make room for TV: Television and the family ideal in postwar america*. Chicago, Uuniversity of Chicago Press.

The 2007 edition of DTT TMT Industry Group's Telecommunications Predictions. Retrieved on January 16, 2007, from http://www.mondaq.com/article.asp?articleid=45596

Zheng, J Y., Wang, X. Applications 4: interactive multimedia systems: Pervasive views: area exploration and guidance using extended image media. In *Proceedings of the 13th annual ACM international conference on Multimedia MULTIMEDIA '05*. New York: ACM Press.

KEY TERMS

Context Awareness: Is a term from computer science that is used for devices that have information about the circumstances under which they operate and can react accordingly. Context aware devices may also try to make assumptions about the user's current situation.

Convergence of Technology: The coming together of two or more disparate technologies. For example, the so-called fax revolution was produced by a convergence of telecommunications technology, optical scanning technology, and printing technology.

ICT: Information technology as defined by the Information Technology Association of America (ITAA) is: "the study, design, development, implementation, support or management of computer-based information systems, particularly software applications and computer hardware." In short, IT deals with the use of electronic computers and computer software to convert, store, protect, process, transmit and retrieve information. Nowadays it has become popular to broaden the term to explicitly include the field of electronic communication so that people tend to use the abbreviation ICT (information and communication technology). Strictly speaking, this name contains some redundancy.

iTV: Interactive TV (iTV) is an umbrella term. Interactive TV is the content and services (in addition to linear TV and radio channels) which are available for digital viewers to navigate through on their TV screen.

Mobile TV: Watching TV on a mobile phone. There are several mobile TV air interfaces competing for prime time. Digital multimedia broadcasting (DMB) is based on the digital audio broadcasting radio standard; digital video broadcast-Handheld (DVB-H) is the mobile version of the international digital TV standard, and forward link only (FLO) is based on QUALCOMM's popular CDMA technology.

Pervasive iTV: An amalgamation between the concepts of iTV and pervasive TV. However this term goes beyond the concept of traditional TV programs data stream and focuses on content personalization and users' creativity, sociability, context awareness, advanced interactivity, immersive environments, convergence (iTV, mobile phones, in-car-navigators and Internet) and connectivity (one to one and one to many).

Pervasive TV: It is an adaptation of the term pervasive computing and it reflects the concept of accessing TV in different contexts such as home, the office, the auto, outdoors thanks to the convergence of technology.

Sociability: Regards the social character of the usage of TV and it involves the identification of suitable applications and interfaces that support social use.

User-Centered Design: UCD is a design philosophy and a process in which the needs, wants, and limitations of the end user of an interface are given extensive attention at each stage of the design process. User-centered design can be characterized as a multi-stage problem solving process that not only requires designers to analyze and foresee how users are likely to use an interface, but to test the validity of their assumptions with regards to user behavior in real world tests with actual users. Such testing is necessary as it is often very difficult for the designers of an interface to understand intuitively what a first-time user of their design experiences, and what each user's learning curve may look like. The chief difference from other interface design philosophies is that user-centered design tries to optimize the user interface around how people can, want, or need to work, rather than forcing the users to change how they work to accommodate the system or function.

UX: User experience (UX) is a term used to describe the overall experience and satisfaction a user has when using a product or system. It most commonly refers to a combination of software and business topics, such as selling over the web, but it applies to any result of interaction design. Interactive voice response systems, for instance, are a frequently mentioned design that can lead to a poor user experience.

ENDNOTES

1 The Virgin Mobile TB trial became a commercial offering in October 2006, but due to lack of interest the service will close in January 2008.

2 Mobility in city centres involves tasks such as commuting, entertainment-seeking, area visiting and dwelling (trips related to shopping or socialising for example) (Hutchinson, 2001).

3 Although Gaver included the postcards within the probes packet, it has been decided to treat them as an autonomous tool in order to extend the number of users involved in the data gathering process.

4 Experiential narratives

5 Glorianna Davenport, Principal Research Associate at the MIT Media Lab in Cereijo Roibas (2003), Ubiquitous media at the intersection: iTV meets Mobile Communications, Panel at the Proceedings of HCI 2003 Conference. Bath.

Chapter III
Problems Rendezvousing:
A Diary Study

Martin Colbert
Kingston University, UK

ABSTRACT

This chapter seeks opportunities to use mobile technology to improve human mobility. To this end, the chapter reports a diary study of university students' use of mobile telephones for rendezvousing—arranging, and traveling to, informal meetings with friends and family. This diary study reveals, and suggests explanations for, a number of deficits in user performance: (1) rendezvousers occasionally become highly stressed and lose valuable opportunities; (2) outcomes are worse when rendezvousing at unfamiliar locations; (3) 31 to 45 year olds report more personal sacrifices than 18 to 30 year olds; and (4) when mobile phones are used on the move, the experience of communication is slightly worse than when phones are used prior to departure. Ways of using mobile technology to make good these deficits are suggested.

INTRODUCTION

Mobile Technology and Human Mobility

Between 1997 and 2001, ownership of GSM mobile telephones rose from 27 percent of the UK public to 73 percent (Butcher, 2003). One important reason for this rapid adoption of mobile technology was anytime, anywhere access to voice telephony. Talking on mobile telephones gave users the freedom to roam away from fixed access points and remain contactable, even when life took them to diverse, unpredictable locations (Palen et al., 2000). Mobile telephony also acted as a flexible, 'proxy' for resources elsewhere. Rather than endure unproductive, 'dead time,' mobile workers could use their mobile phone to have faxes read or sent, or to learn about developments on other projects (Perry et al., 2001). Mobile telephony was also useful for fine-grain, moment-to-moment awareness and co-ordination.

For example, friends out shopping together could split up to visit different shops, and then use their mobile phones to discuss interesting sale items, and arrange how to meet up again (Ling & Yttri, 2001). Mobile technology was adopted, it appears, and in addition to other reasons[1], because it made 'being on the move' less unproductive and smoother-flowing. In this sense, mobile phones improved everyday mobility.

Such improvements are of interest, because everyday mobility is an important activity (Pooley et al., 2005). Mobility has practical value as an activity that enables individuals to function—to eat, work, sleep—and, as such, it is fundamental to society. It also has social value—the movement that makes life possible also enables interactions that support personal relationships, social networks, and local communities. It adds meaning to life and contributes to society, for good or bad. Mobility also acquires meaning itself and so contributes to our definitions of self ("we are how we travel"), and mobility is part of the process by which individuals learn about, and give meaning to, place and space. Finally, mobility has psychological consequences. It encourages individuals to feel a certain way, and to hold certain attitudes.

Future Technology for Mobility and Rendezvousing

Subsequent generations of mobile technology, however, will not necessarily be adopted as widely or as rapidly as GSM phones. To be adopted, broader-band wireless networks, multimedia input and output capabilities, integrated cameras, positioning mechanisms, context sensors, and so on need to be combined into 'packages' of device, service and network that actually improve mobility for many segments of the general public. But what kind of improvement will bring measurable benefits in mobility to users?

Sometimes, potential improvements are relatively easy to identify and confirm. For example, commuting and long distance travel is often "boring." Consequently, the public may want to download music and video files, play computer games, send picture messages, and consume 'live' streams of audio and video data to escape the tedium of waiting rooms, train carriages and other kinds of transit locale (Antilla & Jung, 2006, p. 222). However, other potential improvements are less obvious and less certain. For example, consider rendezvousing, that is, the informal coordination of a face-to-face meeting between friends and family[2]. The shopping rendezvous described in Figure 1 appears "poor" at first glance, because one party arrived late. However, the delay was caused by a traffic jam (and we can not expect mobile IT to free the roads of congestion), and the rendezvousers used existing GSM telephony to adjust their plans and maintain their convenience and comfort. So where is 'the problem'?

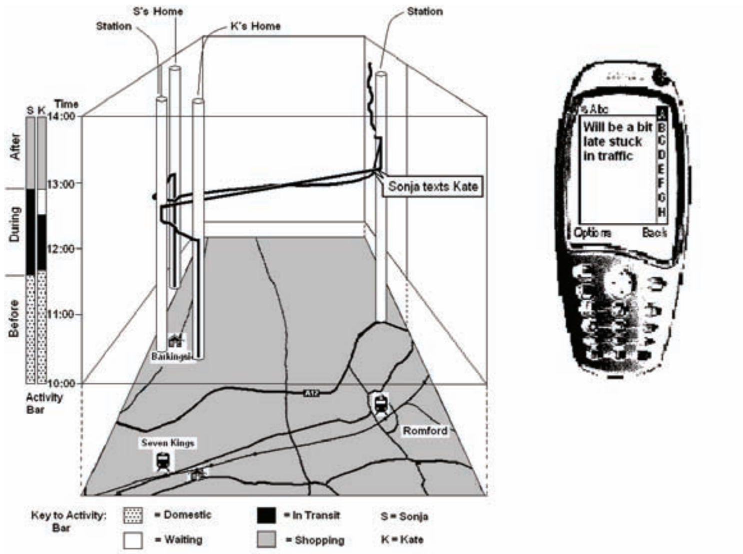
One approach to identifying user problems begins by identifying deficits in user performance, that is, in this case, respects in which human mobility is observably 'worse' under some conditions than others. The identification of any deficit suggests the design goal of removing or "making good" the deficit—a deficit provides a starting point for discussing the improvements that technology might achieve[3].

AIM: TO IDENTIFY OPPORTUNITIES TO IMPROVE RENDEZVOUSING

The work reported here, then, seeks opportunities for mobile technology to improve rendezvousing. To this end, it reports a diary study of university students' use of mobile telephones for rendezvousing. The study reveals a number of deficits in rendezvousing performance, and describes the interactive behaviour that brought about these deficits. Design suggestions that illustrate how the deficits might be made good are then presented[5].

The study extends the literature about user performance with communication systems to cover mobile technology and 'mobile' contexts of use[6]. Previous work concerned table-top communication systems in stationary contexts. This work includes, for example, an investigation of the effects of prepared scripts upon the consensus reached during chat sessions (Farnham et

Figure 1. A rendezvous: Sonja is late meeting Kate to go shopping



al., 2000), a study of the impact of a simplified 'communication computer' upon social interaction, mental stimulation, and attitude of elderly users (Czaja et al., 1993), and a comparison of the effect of video link quality upon overall system ratings and task completion time (Matarazzo & Sellen, 2000).

Participant 32 Entry No. 5 Traffic delays a rendezvous to go shopping. Sonja had arranged to meet an old friend at Romford railway station on Saturday at 12:30 p.m. Then, they would go shopping together. Sonja's friend Kate took the train from her home in Seven Kings, and arrived a few minutes early. Sonja drove from her home in Barkingside, but became stuck in a traffic jam in Romford town centre. At this point, Sonja phoned Kate (using her boyfriend's mobile phone) to say that she would be a little late. Sonja only knew a few landmarks in Romford, but conveyed her position accurately enough. Kate was half expecting traffic delays anyway, and went for a coffee rather than stand in the cold (It was February.). They met up 20 minutes later than planned. This rendezvous is represented in Figure 1[4]

Based upon data collected in 2000 (see Colbert, 2001). All places are in London, UK)

A DIARY STUDY OF RENDEZVOUSING

Study Design and Methodology Issues

Previous studies of communication systems design encourage methods that capture data in the field under naturalistic conditions. 'Practice' studies of communication—first, of landline telephone users (Frohlich et al., 1997; Lacohee & Anderson, 2001) and, more recently, of mobile phone use during urban journeys Tamminen et al., (2003) —and studies of media choice (Kraut et al., 1994; Preece, 2001; Whittaker et al, 1996) —report and illustrate the situated nature of communication. They show how, for example, the answering of a shared, household phone, or the choice to use e-mail rather than have a face-to-face meeting, reflects the situation in which communication occurs. Consequently, studies conducted under controlled conditions may simply distort the communication behavior being studied—participants may change their selection and use of media to better fit the 'experimental' situation. These studies also report and illustrate to the wide range of behavioral and performance parameters that are relevant to communication systems design. For example, the alternative communication services available, the richness of expression desired, the pace of exchange required, and the development of critical mass and social norms all influence usage and preference. These studies indicate, then, that it will be difficult to measure *all* relevant constructs within a single study. Also, any study needs to monitor the use of all the communications media at a users disposal, not just a single medium in isolation.

The work reported here satisfied the requirements for a naturalistic, field study that makes a large number of observations about all communication services on mobile phones[7] by asking participants to keep a diary about actual rendezvous they participated in, and the related communications. Diaries have long been used in user-centred development (Rieman, 1993), particularly to capture data in situ, with minimal observer effects (Carter & Mankoff, 2005). The diary used in this study collected both qualitative and quantitative data about each rendezvous, by having each diary entry comprise of free text, narrative descriptions as well as answers to preset, closed questions, and rating scales.

Method

Participants

The participants in the diary study were 22 male and 22 female students from the School of Computing and Information Systems, Kingston University. The aim of selecting participants was to obtain a sample that was large and gender-balanced from a source that mostly comprised of males. Between January, 2001, and April, 2002, students who took a module in human-computer interaction completed a diary as a minor part of coursework exercises. The diaries concern the participant's rendezvous for one week in January or February. Twenty-two female students completed a diary and consented to its anonymous use here. Then 22 male students were selected from the appropriate module year, to match the female participants as closely as possible in terms of age, ethnic background, marital status, number of children, and mobile phone ownership.

The diary keepers had a mean age of 25 years 11 months. Of the 44 participants, five did not state their age, 30 were in the 18 to 30 year age group, and nine were in the 31to 45 year group. Sixty-six percent were single, 23 percent had been with the same partner for more than one year, and 11 percent were married. Fourteen percent had children. All diary keepers were registered as full-time students, but 33 percent did more than 10 hours per week paid work in addition to their university studies. 89 percent owned a mobile telephone, 89 percent had access to a fixed line telephone, and 98 percent had a private e-mail account in addition to their university account. If they owned a mobile phone, 51 percent used it more than 10 times per week, and, if they had access to a fixed-line phone, 33 percent used that more than 10 times per week. Forty percent of participants were Asian in ethnic origin, 40 percent were European, nine percent were African-Caribbean, five percent were Middle Eastern and seven percent were 'Other'.

Materials

Each diary entry was comprised of: (i) an open-ended, narrative description in the participant's own words of what happened, and why; and (ii) the participant's responses to a questionnaire, which asked for specific details of each rendezvous and associated communication. This questionnaire comprised 37 questions as follows:

- **Questions 1–6:** The event (the who, when, where and why of the events)
- **Questions 7–11:** Outcomes (the additional stress and lost opportunity associated with attempts to meet at the time and place initially agreed)
- **Questions 12–24:** Usage and user experience of communication *prior to departure* for the rendezvous. User experience comprised satisfaction, convenience, social acceptability, disruption, frustration, and mental effort

- **Questions 25–37:** Usage and user experience of communication *whilst en route* to the rendezvous.

Two sets of answers about usage and user experience were returned per rendezvous—one for all communication that occurred prior to departure, and one for all communication that occurred en route. For example, a rendezvouser who spoke once on the phone and sent one text message before departing for the rendezvous point, and then spoke once on the phone and listened to one voice mail en route, provided one combined experience rating for the phone call and text message, and one combined experience rating for the phone call and voice mail. The prior to departure phase ends, and the en route phase begins when the first rendezvouser to do so departs for the rendezvous point.

Procedure

At the outset of the study, all participants were given an overview of future position-aware, computing and communications for mobile devices, and were introduced to the aims of the study and the obligations of diary keeping. To illustrate the kind of services that could be developed, participants examined fixed-access Web sites that provided map, transport, and venue information, such as www.multimap.com, and londontransport.co.uk. To encourage complete, and relevant free text descriptions of events, a possible future service was described, in which each member of a small group were able to display the positions of other group members on their mobile telephone. Participants made one diary entry for each rendezvous event. Participants were encouraged to complete their diary as soon after the event as possible, but were free to choose a time and place that was safe and suitable for thinking and writing. At the end of the diary keeping period, participants summarised their diary and its completeness. Questionnaire responses were processed automatically by an cccular reading Mmachine, which generated a text file that was checked and then read into statistical analysis software.

RESULTS

Rendezvous Reports

The following sections will be illustrated with reference to actual rendezvous reports and quotations from diaries.

Type and Frequency of Rendezvous and Communication

Diary keepers took part in a total of 248 rendezvous—a rate of approximately 5.6 rendezvous per week, or just under one per day. The rendezvous enabled a wide range of subsequent activity—from coffee in the student lounge, and shopping trips, to airport collections and wedding receptions (see Table 1). The rendezvous were very often in locations at which diary keepers had rendezvoused before (65 percent), and included people with whom they had close relationships (close friends 63 percent, immediate family 22 percent, acquaintances 22 percent, extended family 12 percent, and strangers 10 percent). The mean size of rendezvous was 3.6 people, including the diary keeper.

The plan for the rendezvous changed on average 0.56 times prior to departure, and 0.35 times en route—taken together, almost once per event.

Each rendezvous involved, on average, 4.08 communications. About half were telephone calls, and about a quarter were text messages.

The sample of participant activities reflects the relative freedom of student life. Failing to meet as agreed often led to stress and lost opportunity (e.g., P1 #3307), but undesired consequences were not necessarily entailed by being late or changing plans (e.g. lateness can be anticipated and accommodated P42 #3862). Nor was failing to meet as agreed necessary for stress and lost opportunity to arise (e.g. a loose plan was problematic in P33 #4381). Further details of this sample of rendezvous and communications are available in Colbert, 2005a.

Table 1. Example rendezvous

Participant Diary Sheet	Description of Rendezvous (paraphrased for clarity)
P1 #3307	I had arranged for a friend to drive by my house at 3:30 p.m. We would then continue into town and go shopping together. When my friend did not show up, I repeatedly attempted to contact her on the telephone, but each time my friend's phone was engaged. My friend did not reach my house until 4 p.m. Her departure from her house had been delayed because she had been talking to her boyfriend on her mobile phone. By 4 p.m., rush hour had begun, so by the time we got to town, the shop I specifically wanted to visit was closed. I was frustrated, because I wanted to talk to my friend urgently, but her line was busy. If her phone had a call waiting facility (which beeps when another person attempts to telephone the recipient), I may have been able to.
P21 #4340	I had arranged to meet my cousin at the theater. I decided to drive there. However, I was unable to locate the theater. I wanted to telephone my cousin to get information from her, but my mobile phone was not to hand—it was in the bottom of my bag. I was thinking about where it might be rather than the correct location and route to the theater. In the event, I was 10 minutes late.
P25 #1481	Myself and two friends, D and T, arranged to meet in a pub at Waterloo. The afternoon of the meeting, D phoned me saying that now he would probably not be able to make it. I said that he was to call T's mobile phone should he in fact be able to make it, and want to check that we were still in the pub. T kept his phone on the table in the pub, where it could have been stolen.
P33 #4381	I was to meet some friends for a pub crawl in Clapham to celebrate my birthday. No exact time was agreed. Everyone was to arrive at different times, depending upon when they finished work. We selected the pub at which to meet at the last minute, so I contacted my friends to let them know. Some of them had difficulty finding it. I made long phone calls explaining my location. It was very noisy in the pub, so I gave some people directions as text messages, which was very laborious. One person was very late due to "unforeseen complications." Had I been able to let my friends know my current location, it would have spared me the telephone calls and text messages, and we could have moved around more freely.
P39 #4040	I had arranged to meet 25 friends and relatives at my cousin's house, where a convoy of cars would then leave for an engagement in Gloucester. One person was late, so I sent him/her a text message, to discover their whereabouts. This message received no reply, so I sent another message. This message received no reply either. Not getting a reply was frustrating, but I was able to load my car whilst texting. It transpired that the latecomer had received all four messages, but had not replied because he/she was driving. The latecomer reasoned that if he/she stopped the car to answer the phone, it would just make him/her even later.
P41 #1922	I was at work (I serve in a shop) when a friend called me, inviting me over for dinner that evening. When the phone rang, I was serving a customer, so I was speaking on the phone and serving the customer at the same time. I don't see any best way to plan a meeting [other] than talking to the person you supposed to meet. (The rating for social acceptability was only 2 out of 5)
P42 #3862	I was driving to visit my parents in a suburb of North East London when I got stuck in a traffic jam near Kings Cross. I took the opportunity of stationary traffic to warn them I would be late.
P43 #1965	I received a text message inviting me to some drinks that night. They had already started and I was to join them after I finished work. The message read, "Were going to leic sq 4 drinks from 7 onwards, let us know when u there, c ya!" Unfortunately, when I finished work and tried to telephone them, I could get not get a response from them (so frustrating!!), so I left a text message "call me when u get this!" When I received a delivery confirmation for my message, I inferred that they now had reception, so I telephoned them to find out exactly where they were. All that effort for a few words!

Table 2. Frequency of occurrence and mean level of stress and lost opportunity x 'success' of rendezvous

Rendezvous Outcome	Stress Reported (percent)	Level of Stress (mean rating)	Lost Opportunity Reported (percent)	Level of Lost Opportunity (mean rating)
Met as Agreed	16 percent	2.0	12 percent	1.5
Did Not Meet as Agreed	64 percent	1.6	60 percent	1.5

Performance Deficits: Stress and Lost Opportunity

'Severe' Events

When stress or lost opportunity was reported, it was rated as 'medium/high' (4/5) or 'high' (5/5) for 13 percent and 11 percent of rendezvous respectively. This works out as a rate of around one 'severely problematic' rendezvous per month[8]. The most stressful rendezvous' were: an unconfirmed, last minute change of plan for meeting children after school; picking up friends one by one in a van on the way to catch a ferry; a dinner party host who rushed out to look for their dog just as their guests were due to arrive; forgetting to collect a brother from work; and trying to find a relative at the airport when the phone battery was running out. The most valuable opportunities lost arose when: multiple participants arrived late to discuss coursework; the rendezvousers arrived late and lost their reserved table at a restaurant; and, at a music venue, someone who had arrived on time was constantly interrupted from a latecomer who had got lost, and then they had to leave the venue to find him.

Lost opportunities frequently took the form of delay (27 percent). Other types of lost opportunity were reported less frequently—re-structuring—the activity went ahead, but in a different order, with different roles. (14 percent), less participation—the activity went ahead and someone joined in late (12 percent), individual sacrifices—an opportunity to do something for one's own sake was forgone (10 percent), non-participation—the activity went ahead without some individuals)(9 percent), and non-occurrence—the activity was cancelled/aborted (2 percent).

Diary-keepers attributed problems rendezvousing to two causes in particular (see Figure 2—the mode of travel (trains did not run on time, traffic was heavier than expected, etc.), and the over-run of previous activities. Other reasons for rendezvousing problems were cited less frequently—a poor plan (the plan was incomplete, inaccurate, never agreed or forgotten), the failure to value success (someone thought that arriving as agreed was not important), lack of information about other rendezvousers (rendezvousers were not aware that some others were delayed, lost or not coming), and lack of geographic information (rendezvousers become disorientated, or could not find the meeting place). Occasionally, problems were attributed to the performance of additional tasks, and lack of travel information (rendezvousers were unaware of routes, schedules, etc.).

Individual Sacrifices Amongst 31-45 year olds

A comparison of the outcomes reported by participants aged 31-45 years and those aged 18–30 years revealed a number of differences (see Table 3, and Colbert, 2005b):

1. 31-45 year olds more frequently attributed problems to the overrunning of previous-activities, and to taking the opportunity to perform additional, spontaneous tasks ('side-stepping')
2. 31-45 year olds more frequently report that the lost opportunities arising from problematic rendezvous take the form of individual sacrifices

Free text entries in the diaries suggested that these differences arise, because 31-45 year olds have commitments to spouses and children[9], and so pack their program of daily life with planned activities more tightly than 18-30 year olds (Carlstein et al., 1978). If one activity overruns, then it has a knock-on effect upon later rendezvous. There is no slack in the system. ‘Side-stepping’ is seen as a more frequent cause of problems by 31-45 year olds, because side-stepping is a useful technique for “getting everything done.” It increases the proportion of time spent being “productive” relative to time spent travelling (which is “unproductive”). 31-45 year olds are more likely to perceive lost opportunities in the form of personal sacrifices, because 31-45 year olds are more aware of the activities they could have packed in to the time they actually ‘wasted’ failing to meet as initially agreed. It is as if 31-45 year olds have lengthy ‘to do lists’ (Taylor & Swan, 2004) continually at the back of their minds—tasks which they would like to perform, if only they could find the time.

Stress and Lost Opportunity at Unfamiliar Rendezvous Points

A comparison of rendezvous at familiar and unfamiliar meeting points revealed a number of differences. When meeting at unfamiliar places (places at which rendezvousers had *not* met before), the diary keeper:

- Reported stress more frequently
- Reported higher levels of stress and lost opportunity (see Table 4)

The reason for these performance deficits is as expected. When meeting at unfamiliar locations, rendezvousers more frequently attributed problems to lack of geographic and travel information (p<0.001 and p=0.011 respectively)(see Figure 3, and Participant 21 event #4340, and P33 #4381 in Table 1). The lack of this kind of information underlay the apparent, but not statistically significant, increase in related reasons for problems—the mode of travel, and poor planning (P43 #1965). It is possible, however, that the deficits noted may also be due to the fact that meetings at unfamiliar locations more often included strangers (Colbert, 2004, p=0.013). For example, a dinner party may occur at an unfamiliar location, because the diary keeper knows other guests, not the host. Meetings with strangers are slightly more formal, because “first impressions count,” so stress and lost opportunity ratings increase, as rendezvousers are more sensitive to “being late.”

Figure 2. Reasons for failing to meet as initially agreed (%)

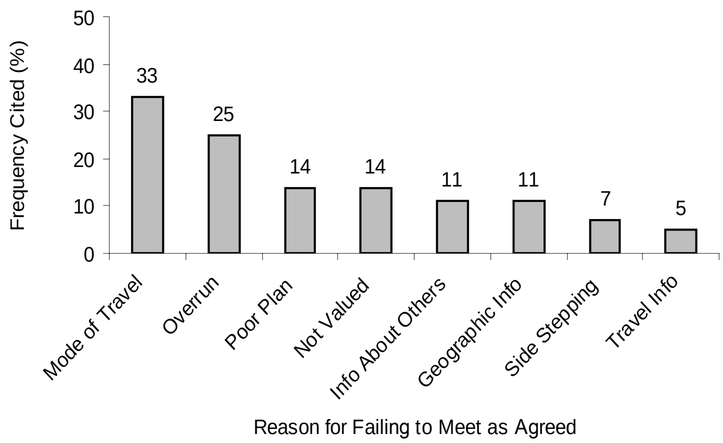


Table 3. Rendezvousing outcomes: Age differences

Measure	% 18-30s	% 31-45s	Sig.
lost opportunities take the form of personal sacrifices	7%	22%	p = 0.008
problems attributed to overrunning of previous activity	21%	39%	p = 0.050
problems attributed to 'side step-ping'	5%	17%	p = 0.002

Performance Deficits: User Experience

A comparison of user experience of communication (phone use) in different contexts—prior to departure and en route to the rendezvous point—reveals that, when used en route, many aspects of user experience of communication are impaired by a small amount (Colbert, 2005b). Communication en route is significantly more frustrating, less convenient, more disruptive and less socially acceptable than communication prior to departure (see Table 5). These findings are strongly significant. At least the 0.001 level, using a paired-samples two-tailed T-test. Communication is also less satisfying, although this finding is only significant at the 0.05 level. Note, however, that the size of the impairment on each scale is not great (around a third of a rating point on a five point scale).

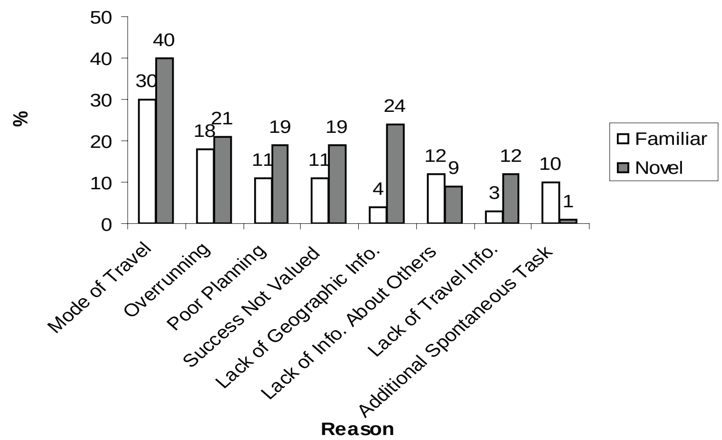
Free text entries in diaries suggest that these impairments are due to the cumulative effect of various adverse factors that tend to be more common and more severe in a 'mobile' context of use. These factors were:

- **Lack of network coverage:** Rendezvousers underground (e.g., in car parks, or tunnels), in the signal 'shadows' cast by tall buildings, or suffering interference from other activity in the airspace, may suffer low quality connections, be unable to connect, or 'cut off' in mid conversation (see P43 #1965, P2 #3164, P6 #0290 in Table 6).
- **'Phone free' zones:** Rendezvousers are asked to turn off their mobile phones in many public places, either for the sake of bystanders (as in 'quiet' railway carriages, theaters and some restaurants) or for reasons of safety (e.g., in hospitals, teaching laboratories and aeroplanes). P26 #1504 and P39 #4040 (see Table 6) were in a phone free zone, because holding a phone whilst driving is illegal in the UK.
- **Environmental noise:** Busy streets, and noisy vehicles or train stations, or entertainment venues sometimes make use of the telephone unpleasant or impossible, and tend to keep the duration of calls to a minimum (see P 26 #1506, P34 #4284, and P36 #4003, Table 6).

Table 4. Rendezvousing outcomes: Familiar vs. unfamiliar locations

Measure	Unfamiliar	Familiar	Significance
% reported stress	61%	45%	p=0.023
Stress (mean rating 1= low; 5=high)	2.54	2.13	p=0.04
Lost Opportunity (mean rating 1= low; 5=high)	2.26	1.87	p=0.04

Figure 3. Reasons for not meeting as initially agreed: Unfamiliar vs. familiar rendezvous points



- **User availability:** The failure to be available for urgent communications sometimes reduces satisfaction and acceptability (see P1 #3307). Conversely, making oneself available to communicate whilst engaged in another activity sometimes feels disruptive (see P26 #1504 and P41 #1922).
- **Workspace constraints:** The space immediately around a rendezvouser is organised to ease the task of transiting to the rendezvous point, or to suit the locale, rather than communicating (see P21 #4340).
- **Security:** Using a mobile phone in some contexts sometimes feels 'unsafe,' because it might be lost or stolen (see P25 #1481).
- **Codes of conduct for public spaces:** The requirements of communication sometimes conflict with informal expectations for behaviour in transit locales (see P6 #0290).
- **Time pressure:** Time pressure sometimes impairs many aspects of user experience. For example, one diary keeper reported sending an e-mail to a group of friends and acquaintances, despite knowing that one recipient would not check her e-mail account soon enough for the message to be useful. However, the invitation had to be distributed rapidly, so he sent the e-mail anyway, albeit at the cost of excluding his friend. He felt this was not really acceptable, but he had to send the message immediately. Time pressure is also implicated in P1 #3307, P21 #4340, P39 #4040 and P43 #1965.
- **Lack of device power:** If a phone battery is running low, it is frustrating when the owner wishes to make a call, or inconvenient if he or she has to borrow a power adapter and recharge at university. One rendezvouser, whose batteries were on the verge of expiring, reported sending the briefest text message, when they would have preferred to telephone.
- **Conflict with preferred life-paths:** Rendezvousers have programmes of personal activities—plans that specify how they intend to use their limited resources of time and space. The need to communicate sometimes interferes with these plans, particularly if communication is constrained by lack of network coverage and phone free zones, limited user availability and time pressure. Taking detours, or delaying travel, in order to communicate is sometimes frustrating and inconvenient. The latecomer in P39 #4040 and P42 #3862 delayed communication until it did not interfere with the preferred life-path.

The impairment in user experience en route to a rendezvous, may also be due to rendezvousers'

Table 5. User experience of communication: Prior to departure vs. en route

User Experience	Prior to Departure (rating)	En Route (rating)	Significance
Satisfaction	3.92	3.75	p=0.036
Convenience	4.10	3.77	p=0.001
Social Acceptability	4.31	4.00	p=0.001
Frustration	1.42	1.69	p=0.001
Disruption	1.35	1.63	p=0.001

Table 6. Selected quotes from diary entries

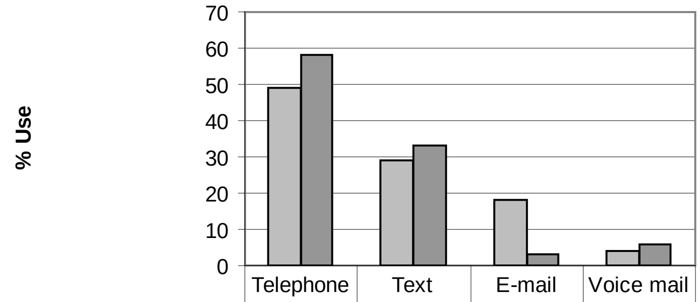
Participant Diary Sheet	Quote
P2 #3164	"I also used text messaging when my cousin was on the train, as she did not have any reception to pick up the phone. . .bv.less agitated and frustrating. When in a distressing situation, or a situation that may cause anxiety, having a device that allows you to contact a person straight away can provide some relief."
P6 #0290	"During the rendezvous the mobile was slightly less satisfying to use as the reception was not very good on my travels. It was also slightly disruptive to the other passengers when I was on the train."
P26 #1504	"Driving whilst talking on the phone is not only illegal but hard! This forced me to pull over a couple of times and make a call to my girlfriend to get some more directions."
P26 #1506	"In the midst of a pub atmosphere i.e. where there is a lot of people talking, shouting and loud music, hearing your mobile phone ring is virtually impossible. There fore have a voice mail service that tracks your calls takes messages for you and then returns the messages to you is not only satisfying but reassuring."
P34 #4284	"My friend was unable to use his mobile phone because of the noise in side. He had to go outside to make the call."
P36 #4003	"It was annoying that I did not hear the phone beep to indicate a message. The message was from here friend who was late collecting me from my house en route to a dinner party"

tendancies to use the telephone more (p=0.021), and e-mail less (p<0.001), when en route (see Figure 4). En route to a rendezvous point, rendezvousers need to use the channel of communication that grounds information almost instantly (the telephone). Awkwardly, the telephone is the communication medium whose user experience is most impaired by the context of being en route.

DESIGN IMPLICATIONS

These performance deficits suggest the goals of developing applications of mobile technology that make good these deficits. The following paragraphs consider each deficit identified in turn, and illustrate the kind of mobile technology that could conceivably respond to it[10]. When deciding whether or not to respond to these deficits, and if so, then how, it is first worth noting the importance and general relevance of rendezvousing. Rendezvousing is important for mobile technology design, because around one third of current mobile phone use is rendezvous-related. One study estimated that 34 percent of telephone calls made from mobile phones by working parents were travel-related (Ling & Haddon, 2001). In Grinter and Eldridge's study, 36 percent of text messages sent by teenagers were related to coordination with friends or family, and coordination

Figure 4. Usage of communication services prior to departure vs. en route



	Telephone	Text	E-mail	Voice mail
Prior to Departure	49	29	18	4
En Route	58	33	3	6

Communication Service

was also found to be the most common topic for text messages sent by Norwegian users (Telenor, 2002 cited in Ling, 2004). Rendezvousing is also relevant to a wide range of potential applications, including group-based communication and awareness systems (Milewksi & Smith, 2000; Nardi et al., 2000; Tang et al., 2001), electronic guides (Chincholle et al., 2002; Poposchil et al., 2002), position-aware reminder services (Marmasse & Schmandt, 2000), and diaries and event planners (Pousman et al., 2003).

Responding to Occasional High Stress and Lost Opportunity: Controlled Disclosure of Position Integrated with Group Communication

Assuming that one 'severely problematic' rendezvous per month is sufficient to limit individual mobility, position-aware, or location-enhanced communication for small groups appears potentially useful as a complement to telephony and text messaging. At face value, position and location[11] information is relevant to many of the frequently cited reasons for failing to meet as agreed (see Figure 2), for example, by letting rendezvousing 'A' decide whether that rendezvouser 'B' still intends to participate, and if so, then by how long is he or she delayed, or that 'B' has chosen to not come and do something else. The opportunity to obtain information about a number of other rendezvousers, without interrupting them, and without a lot of effort oneself, also appears to have some advantages (see the problems of scale P33 #4381 and 'no response' P39 #4040). However, the need to communicate via phone or text still remains, for example, to confirm inferences, in addition to negotiating changes of plan, and providing information.

A recent Wizard of Oz study asked participants to execute 'contrived' rendezvous with and without a mobile phone based 'friend finder' service in a local urban area. The study supports the potential efficacy of such systems for rendezvousing. Participants rapidly perceived the potential utility of the simulated service, and often used tracking in certain situations, for example, to discover context information before making a follow-up telephone call, if needed (Dearden et al., 2005).

The obvious limitation of tracking services, however, is that inappropriate use may invade other users' privacy. Preparatory studies suggest users will permit others to access the location details they require, provided they are satisfied with who wants to know, and why they need to know it (Barkhuus & Dey, 2003; Consolvo et al., 2005). Another study, partly in view of privacy issues, suggests an alternative to full integration of communication and tracking—simply broadcasting one's own location as an invitation to socialise

(Farnham, 2006). This study deployed SWARM, a service which broadcast short messages to a small group of extremely social, urban professionals. The broadcasts were frequently used for rendezvousing, notably for 'fishing' for company, or 'scouting' to find the best event. However, broadcasts could also make other contributions to rendezvousing—notably, announcing a delay or a decision not to attend. It will be interesting to see whether 'tracking' or 'broadcasting' position and location is adopted most widely, and whether either service reduces the frequency of 'severely problematic' rendezvous. Important questions for future work are, "Does the avoidance of occasional 'severely problematic' rendezvous constitute a 'need to know' (and so granting tracking rights)?" and "Is the avoidance of occasional 'severely problematic' rendezvous sufficient motivation to seek the permission to track?" Just encouraging others to broadcast their position, if they think it is necessary, might be enough.

Responding to More Individual Sacrifices by 31-45s: Reminder Systems

Reminder systems are relevant to rendezvousing because reminding prepares rendezvousers for side-stepping should the opportunity arise en route to a rendezvous. Reminders may also prove more useful than 'locator' services (e.g., "Where is the nearest … <e.g., bank>?" because the majority of rendezvous occurred at familiar places, so rendezvousers probably know about the availability of relevant resources, they just fail to recall their need to make a visit when the opportunity to 'call in' arises.

The 'position-aware reminder' application is suggested in Figure 5, on the assumption that an effective reminder occurs when and where a user has the opportunity to act upon it, not where and when the user recalls the need to do something (see Colbert, 2004). This application seems most likely to be useful, when the rendezvous, and the route to it, are regular, say, collecting the kids from school, going to the gym, commuting home etc.

Interestingly, a recent Finnish field trial of 'De-De,' a context-enhanced phone, reports that users sometimes used a location-aware text messaging facility based upon cellID to deliver 'prompts' to friends to do something (Jung et al., 2005). Although De-De could have been used this way, participants did not send reminder messages to themselves to help them side-step, presumably because the participants were teenagers (so their life-situations did not warrant reminders to themselves), and because this usage was not prominently articulated in the user interface, as it would have been, for example, with a location enhanced to do list, or personal calendar.

An informal survey of another 20 students at Kingston University suggests that reminders near the shop, are not always the best possible reminders, because the minder comes too late. The time for the rendezvous is now so close that there is no time to take a detour. The best possible reminder, and particularly the widest acceptable reminder, was sometimes triggered by departure for the rendezvous point i.e. leaving campus, or leaving home. Kim recently conducted an exploratory user study with 'Gate Reminder' (Kim et al., 2004). Located at the front door of a family home, Gate Reminder detects RFID tags on participants and objects as they cross the threshold, and displays visual reminders on screen. Although the form of interaction was satisfactory to many participants, the additional infrastructure was felt to be too 'heavy weight' given the benefits. A more recent field study of PlaceMail (a system similar to De-De but augmented with GPS and with voice input) confirmed that users often found locational reminders to be useful (33 percent, Ludford et al., 2006, p896). Also, if the reminder is given at the "right place" according to the user, then the reminder is more likely to be seen as useful and acted upon.

From a rendezvousing point of view, future work needs to pay close attention the effect of such location-based reminders upon user performance and communication. In effect, their use in effect enables 31-45s to delay planning their day, and to pack their daily schedule even tighter with activities. This may create even greater demands for micro-coordination—and so greater use of the mobile phone for communication, more frequent changes of plan and more frequent 'failure' to meet as initially agreed.

Responding to Worse Outcomes when Meeting at Unfamiliar Locations: Personal Route Planning, Navigation and Information Seeking

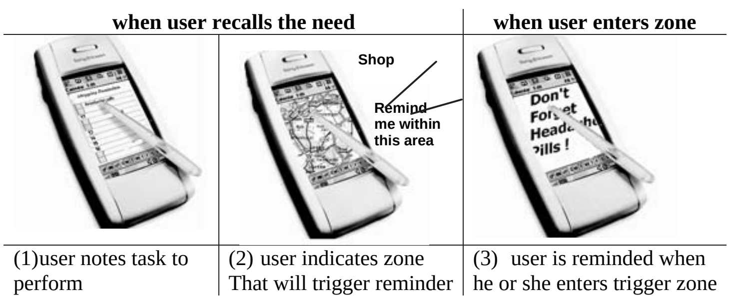
It is not an original to suggest that future mobile technology could provide pedestrians, cyclists, and those on public transport with the same level of support for orientation, route planning and local information seeking that is already available to motorists. The distinctive requirements of pedestrians, for example, the usefulness of shop signs as landmarks for navigation, have already been investigated (May et al., 2003). It is probably reassuring to those already implementing such systems that there is, indeed, a problem to solve. However, many systems under development target tourists. This study reminds us that 'locals,' too, sometimes visit new parts of their own city, and perhaps locals have distinctive requirements. For example, they may be more likely to be listening to music whilst traveling, and so favor systems that provide navigational guidance through this audio channel (e.g., Warren, 2005). It also reminds us that rendezvousers want to work with spatial information at all phases of the activity—both prior to departure, and en route (and there may be many stages to their journey). In this light, integrated services that coordinate use of specialised desk-top, in-car, and personal devices (e.g., Baus et al., 2002) seem to have potential.

Responding to Impaired Experience en Route: Better Connectivity, Faster-Paced Messaging, Awareness and Negotiation, Systems Fit-For-Contexts

Better Connectivity

Given the wish the find an application for broader-band, 3G networks, it is a pity that this study suggests that better connectivity i.e., more complete coverage and constantly high quality of service, with existing 'second generation' (2G) networks will improve user experience of communication en route to a rendezvous, more than additional bandwidth (cf. the frustration and disruption caused by incomplete network coverage and poor quality of service.) The financial costs of 3G networks are such that they may never extend far beyond densely populated areas. A more promising development for rendezvousers is the provision of wireless local area network (LAN) access in public places (shopping centres, railway companies, etc.). Rendezvousers needing to communicate, but beyond the reach of a 2G network, may connect to a wireless LAN and, from there, to the world at large. Alas, it is not until the so-called Fourth Generation (4G) of wireless network that the

Figure 5. A location-enhanced to do list for mobile phones



seamless integration of personal, local and wide area networks and ''multi-mode'' mobile phones are envisaged. Even then, the level of connectivity actually achieved with 4G remains to be seen.

Faster-Paced Messaging

This study also suggests that 'faster-paced' text messaging may also improve user experience of communication during rendezvous. Currently, rendezvousers tend to use the telephone during rendezvous rather than text messaging, because text messages are exchanged and grounded too slowly. Telephone calls are made despite the fact that the experience of telephony appears to be more impaired by the context of 'being en route' than texting. With a 'recorded delivery express text' service, selected messages receive priority transmission when networks are congested, so senders can be guaranteed a time of delivery, say, within 30 seconds of sending, provided that the recipient's device is on, and within coverage. This service would also acknowledge: (1) transmission from the sender (many networks/phones already provide this); (2) transmission to the recipient (some networks/phones provide this); and (3) reading by the recipient (not currently provided for text, but provided for e-mail). Another alternative resembles a 'response paid text' service. With this service, recipients are asked to select one of a limited number of responses to a message, and the sender pays for the response. For example, the message 'running 15 minutes late' may require the response 'OK,' 'Not OK,'or 'Ask again later.'

Awareness and Negotiation

Experience ratings may also be improved by providing users with means to negotiate appropriate contexts in which to communicate, and appropriate communication media to use in this context. Recent work has focused upon 'awareness systems' (Tang et al., 2001; Milewski and Smith, 2000), which inform users initiating communication about the other party's current, or previous activity. Relevant activity information may include a user's location, the status of a device (on/off), service usage, the other party's speed and direction, mode of transport and personal calendar. Devices and services could also provide users with greater control and pre-programming of their 'reachability.' For example, some users in this study turned their phone off to prevent the receipt of telephone calls whilst driving—a somewhat drastic action, as it also prevents receiving text messages. User experience may be improved if users could temporarily, or locally, 'bar' selected media (say, the telephone), but accept others (say, text messages, e-mail or voice messages). These preferences may be controlled by setting the device's 'mode'. However, they could also be pre-programmed in advance. Some users in this study took regular routes and stated, for example, that they preferred not to use the phone in certain contexts, such as crossing a busy road junction, or boarding their bus. On fixed networks, instant messaging is often used to negotiate good times to talk or to e-mail. If provided on mobile devices, it could serve a similar purpose for rendezvousers.

Devices Fit-for-Context

It is particularly important to improve the experience of telephone calls en route to a rendezvous point—these calls accounted for almost 60 percent of communications. Recent innovations, such as caller ID, which makes it easier for rendezvousers to ensure that they accept urgent calls, but leave non-urgent calls for a more suitable context, may already provide some help. Other innovations, such as mechanisms to cope with environmental noise (pre-recorded "Go on. I'm still listening messages," whisper mode, Nelson et al., 2001), and wearable devices (Sawhney & Schmandt, 2000) may also increase satisfaction and convenience. A possible limitation of such approaches, however, is that synchronous, real-time services such as the telephone require both participants to be in a suitable context at the same time to achieve good user experience. Such a situation may be difficult to achieve. Also, even if devices are more resilient to some adverse factors, such as environmental noise, synchronous audio services may still be impaired by other factors (phone-free zones, user non-availability, parallel task performance, etc.).

FUTURE WORK

This chapter has sought out opportunities to improve one kind of human mobility by identifying some deficits in current performance. Having identified these opportunities, it is logical for future work to make good those deficits by implementing and testing systems of the kind outlined in section 5. Quite independently of this work, many such systems, and other kinds too, are already under development.

Of course, various activities and information influence the design goals that are set for specific development projects—not just performance deficits from user studies, but user requirements from focus groups, feature lists from market research, and so on. The advantage of statements of current user performance obtained by this study, however, is that they provide a baseline against which levels of performance achieved by future users and systems can be compared. For example, one could imagine repeating the study reported here in say, 2012 (by which time future technology will have been widely adopted by student participants). How will context-aware communication for small groups, personal navigation, and reminding have affected rendezvousing performance? And will certain kinds of rendezvous become more popular, as users realise technology enables them to execute more fluid arrangements. It depends on the quality of the design and implementation. So let us hope this goes well.

It will be interesting to see how future devices and services benefit the occasional, highly problematic rendezvous that is the target of redesign. It will also be interesting or perhaps increase the popularity of certain kinds of rendezvous—flash mobs, en route, delay planning in detail until the last minute.

It is also worth recalling that this study concerned only one kind of everyday mobility. Mobile technology also has the potential to benefit, say, commuting, or long distance business travel. If we identified performance deficits for these kinds of mobility too, then we could similarly inform design goals for other kinds of mobile application (see Colbert & Livingstone, 2006 for an initial attempt).

REFERENCES

Anttila, A., & Jung, Y. (2006). Discovering design drivers for mobile media solutions. In *CHI 2006 Extended Abstracts on Human Factors in Computing Systems* (pp. 219-224). New York: ACM Press.

Barkhuus, L., & Dey, A. (2003). Location-based services for mobile telephony: A study of users' privacy concerns. In *Proceedings of Interact 2003* (pp. 709-712). Berlin: Springer Verlag.

Baus, J., Kruger, J., & Wahlster, W. (2002). A Resource-adaptive mobile navigation system. In *Proceedings of Intelligent User Interfaces 2002* (pp. 15-22). New York: ACM Press.

Butcher, M. (2003). Getting the cold shoulder. *The Guardian newspaper*, April 23rd, 2003. Retrieved January 19, 2007 from http://shopping.guardian.co.uk/gadgets/story/0,,928162,00.html

Cadiz, J. J., Czerwinski, M., McCrickard, S., & Stasko, J. (2003). Providing elegant peripheral awareness. In *CHI 2003 extended abstracts on human factors in computer systems* (pp. 1066-1067). New York: ACM Press.

Carlstein, T., Parkes, D., & Thrift, N. (1978). *Human activity and time geography: Timing space and spacing time (Vol 2).* London: Edward Arnold.

Carter, S., & Mankoff, J. (2005). When participants do the capturing: the role of media in diary studies. *Proceedings of CHI 2005 Human Factors in Computing Systems.* (pp. 899-908). New York: ACM Press.

Chincholle, D., Goldstein, M., Nyberg, M., & Eriksson, M. (2002). Lost or found? A usability evaluation of a mobile navigation and location-based service, in human-computer interaction with mobile devices. In *Proceedings of Mobile HCI 2002* (pp. 211 – 224). Berlin: Springer-Verlag.

Colbert, M. (2001). A diary study of rendezvousing: implications for position-aware computing and communications for the general public. *Pro-*

ceedings of ACM GROUP'01 (pp. 15-23). New York: ACM Press.

Colbert, M. (2004). Rendezvousing at familiar and unfamiliar places. *Journal of Navigation, 57*(3), 327-338.

Colbert, M. (2005a). User experience of communication before and during rendezvous. *Behaviour and Information Technology, 24*(6), 449-470.

Colbert, M. (2005b). Age differences rendezvousing: reminders for side-stepping. *Personal and Ubiquitous Computing, 9*(6), 404-412.

Colbert, M., & Livingstone, D. (2006). Context changes important for talking and text messaging during homeward commutes. *Behaviour and Information Technology, 25(5),* 433-442.

Consolvo, S., Smith, I. E., Matthews, T., LaMarca, A., Tabert, J., & Powledge, P. (2005). Location disclosure to social relations: why, when, & what people want to share. In *Proceedings of CHI 2005 Human Factors in Computing Systems* (pp. 81-90). New York: ACM Press).

Czaja, S J., Guerrier, J. H., Nair, S. N., & Landauer, T. K.(1993). Computer communication as an aid to independence for older adults. *Behaviour and Information Technology, 12*(4), 197-207.

Dearden, D., Hawkey, K., & Inkpen, K. M. (2005). Rendezvousing with location-aware devices: enhancing social coordination. *Interacting with Computers, 17,* 542–566.

Farnham, S.D., & Keyani, P. (2006). Party on: Hyper awareness, hyper coordination, and smart convergence through mobile group text messaging. In *Proceedings of HICSS-39,* London: IEEE Computer Society.

Farnham, S., Chesley, H. R., McGhee, D. E., Kawal, R., & Landau, J. (2000). Structured onlineinteractions: improving the decision-making of small discussion Groups. In *Proceedings of CSCW 2000* (pp. 299-308). New York: ACM Press.

Frohlich, D.M., Chilton, K. & Drew, P. (1997). Remote homeplace communication: what is it like and how might we support it? In *People and Computers XII, Proceedings of HCI'97* (pp. 133-153) London: Springer-Verlag.

Gray W, D., John B. E., & Atwood M. E. (1993). Project Ernestine: Validating a GOMS analysis for predicting and explaining real-world task performance. *Human Computer Interaction,* 8, 237-309.

Grinter, R. E., & Eldridge, M. A. (2001). Y do tngrs luv 2 txt msg? *Proceedings of ECSCW 2001.* Bonn, Germany: Kluwer.

Jung, Y., Persson, P., & Blom, J. (2005). DeDe: Design and evaluation of a context-enhanced mobile messaging system. In *Proceedings of CHI 2005 Human Factors in Computing Systems* (pp. 351-360). New York: ACM Press.

Kim, S. W., Kim, M. C., Park, S. H., Jin, Y. K., & Choi, W.S. (2004). Gate reminder: A design case of a smart reminder. In *Proceedings of Designing Interactive Systems 2004* (pp. 81-90). New York: ACM Press.

Kraut, R. E., Cool, C., Rice, R. E., & Fish, R. S. (1994). Life and death of new technology: task, utility and social influences on the use of a communication medium from video phoning to video interacting. In *Proceedings of CSCW'94 Conference on Computer-Supported Cooperative Work* (pp.13-21). New York: ACM Press.

Lacohee, H., & Anderson, B. (2001). Interacting with the telephone. *International Journal of Human-Computer Studies, 54*(5), 665-699.

Ling, R. (2004). *The mobile connection: The cell phone's impact on society*. New York: Morgan Kaufmann.

Ling, R., & Haddon, L. (2001). Mobile telephony, mobility and the coordination of everyday life. paper presented at 'Machines that become us' conference, Rutgers

University, April 18-19, 2001. Retrived December 8, 2006 from http://www.richardling.com/papers/2001_mobility_and_the_coordination_of_everyday_life.pdf

Ling, R., & Yttri, B. (2001). Hyper-coordination via mobile phones in Norway. In J. Katz and M. Aakhus (Eds.) *Perpetual contact: mobile communication, private talk, public performance*. (pp. 139–169). Cambridge, UK: Cambridge.

Ludford , P. J., Frankowski , D., Reily , K., Wilms, K., & Terveen, L. (2006). Because I carry my cell phone anyway: functional location-based reminder applications. In *Proceedings of CHI2006, Human Factors in Computing Systems* (pp. 889-898). New York: ACM Press.

May, A. J., Ross, T., Bayer, S. H. & Tarkiainen, M. J. (2003). Pedestrian navigation aids: Information requirements and design implications, *Personal and Ubiquitous Computing, 7*(6), 331-338.

Marmasse, N., & Schmandt, C. (2000). Location-aware information delivery with comMotion. *Intl. Symposium on Handheld and Ubiquitous Computing* (pp. 64-73).

Matarazzo, G., & Sellen, A, (2000). The value of video in work at a distance: Addition or distraction? *Behaviour and Information Technology, 5*(19), 339-348.

Milewski, A. E., & Smith, T. M. (2000). Providing presence cues to telephone users. In *Proceedings of CSCW 2000 Conference on Computer-Supported Cooperative Work* (pp. 89–96). New York: ACM Press.

Nardi, B., Whittaker, S., & Bradner, E., (2000). Interaction and outeraction: Instant messaging in action. In *Proceedings of CSCW 2000 Conference on Computer-Supported Cooperative Work* (pp. 79–88). New York: ACM Press.

Nelson, L., Bly, S., & Skoler, T. (2001). Quiet calls: talking silently on mobile phones speech studies. In *Proceedings of CHI 2001 Conference on Human Factors in Computing Systems* (pp. 174-181). New York: ACM Press.

Newman, W.M., & Taylor, A. (1999). Towards a methodology employing critical parameters to deliver performance improvements in interactive systems. In *Proceedings of INTERACT'99* (pp. 605–612). Amsterdam: IOS Press.

Newman, W. M., Taylor, A. S., Dance, C. R. & Taylor, S. A. (2000). Performance targets, models and innovation in interactive system design. In *Proceedings of Designing Interactive Systems* (pp. 381-387). New York: ACM Press.

Palen L., Salzman, M., & Youngs, E. (2000). Going wireless: behaviour and practice of new mobile phone users. In *Proceedings of CSCW 2000 Conference on Computer-Supported Cooperative Work* (pp. 201–210). New York: ACM Press.

Perry, M., O'Hara, K., Sellen, A., Brown, B., & Harper, R. (2001). Dealing with mobility: understanding access anytime, anywhere. *ACM Transactions on Computer -Human Interaction, 8*(4), 323–347.

Pooley, C., Turnbull, J., & Adams, M. (2005). "... everywhere she went I had to tag along beside her": Family, life course and everyday mobility in England since the 1940s. *The History of the Family: An International Quarterly, 10*(2), 119-136

Poposchil, G., Umlauf, M. & Michlmayr, E. (2002). Designing Lol@, a mobile tourist guide for UMTS. In *Proceedings of Mobile HCI 2002* (pp. 155–169). Berlin: Springer-Verlag.

Pousman, Z., Iachello, G., Fithian, R., Moghazy, J., & Stasko, J. (2004). Design iterations for a location-aware event planner. *Personal Ubiquitous Computing, 8*(2), 117-125.

Preece, J. (2001). *Online communities: Designing usability, supporting sociability*. New York: Wiley.

Rieman, J. (1993). The diary study: a workplace-oriented research tool to guide laboratory efforts. In *Proceedings of INTERCHI'93* (pp. 321-326). Amsterdam: IOS Press.

Sawhney, N., & Schmandt, C. (2000). Nomadic radio: speech and audio interaction for contextual messaging in nomadic environments. *ACM Transactions on Computer-Human Interaction, 7*(3), 353-383.

Tamminen, S., Oulasvirta, A., Toiskalio, K., & Kankainen, A. (2003). Understanding Mobile

Contexts. In *Proceedings of Mobile HCI 2003* (pp. 135-143). Berlin: Springer-Verlag.

Tang, T., Yankelovich, N., Begole, J., Van Kleek, M., Li, F., & Bhalodia, J. (2001). ConNexus to Awarenex: cxtending awareness to mobile users. In *Proceedings of CHI2001 Conference on Human Factors in Computing Systems* (pp. 221-228). New York: ACM Press.

Taylor, A. S., & Swan, L. (2004). List making in the home. In *Proceedings of CSCW04 Computer Supported Collaborative Work* (pp. 542-545). New York: ACM Press.

Warren, N., Jones, M., Jones, S., & Bainbridge, B. (2005). Navigation via continously adapted music. In *Extended Abstracts on CHI 2005 Conference on Human Factors in Computing Systems* (pp.1849-1852). New York: ACM Press.

Whittaker, T. (1996). Talking to strangers: an evaluation of the factors affecting electronic collaboration. In *Proceedings of CSCW'96, Computer Supported Collaborative Work* (pp. 409-418). New York: ACM.

KEY TERMS

After the Rendezvous: From the time at which the last rendezvouser arrives at the rendezvous point.

En Route: From the time at which the first rendezvouser to depart does so, until the last rendezvouser arrives at the rendezvous point.

Life-Path Diagram: A continuous representation of human activity in time and space, and with respect to other entities and features of the environment.

Lost Opportunity: What the rendezvouser would have done, had the rendezvous occurred as intended, but which is now impossible.

Prior to Departure: Until the time at which the first rendezvouser to depart does so.

Rendezvousing: Everyday coordination, 'meeting up' of friends and family. The process of arranging, and traveling to a rendezvous point, in order to pursue some non-work, group activity, for example, to watch a movie, or to have lunch. Rendezvousers have personal relationships with each other—they are not impersonal embodiments of organisational roles. So rendezvousing does not include formal or anonymous attendance at institutions, such as 'reporting to the tax office for interview,' 'going to my electronics lecture,' or 'going to the annual general meeting.' It also does not include receipts of service, such as 'having a pizza delivered.'

Time Geography: A school of human geography that emphasises the development and use of continuous models of human activity with respect to time and space, and argues that these models are a basic component of the understanding of spatial behaviour. Classical, 'spatial' geography orients towards space over time, and concerns, for example, changes in settlement *size* and *layout*. Time geography, in contrast, orients towards the activities of human individuals in the context of time space, and concerns, for example, patterns of *commuting* and *migration* into and out of a settlement.

User Performance: Effectiveness of human-computer interaction. The quality of task outcomes achieved, for the costs that users incur achieving these outcomes.

ENDNOTES

1. Mobile technology was adopted for many reasons. See, for example, the communicative opportunities that text messaging provided teenagers (Grinter & Eldridge, 2001).
2. Rendezvousing is an important form of mobility, because it enables actual presence and participation in different social groups. The more and better we rendezvous, the less we have to 'tag along' with, or 'drag around' a group that temporarily does not interest us (Pooley et al., 2005). We are free to disperse,

because we are confident we can successfully get back together again, and that we will stay safe whilst apart.

3 User performance, here, refers to the quality of task outcomes achieved, for the costs that users incur achieving these outcomes. A performance-oriented approach is characteristic of an engineering approach in many domains (Newman & Taylor, 1999). Performance parameters have been productively applied to the design of telephone operator workstations (Gray et al., 1993), and video-based, document capture tools (Newman et al., 2000). They also seem applicable to peripheral notifications (Cadiz et al., 2003), so a performance-oriented approach to mobile technology and everyday mobility would be expected to be productive also.

4 Figure 1 is a life-path diagram—a kind of diagram devised by time geographers to represent human mobility. In life-path diagrams, two dimensions of geographic space (longitude and latitude) are represented 'horizontally' as an apparent surface (in this case, the map of East London suburbs at the foot of Figure 1). Time is represented along a vertical axis, and so successive snapshots of geographic space combine to form an apparent column of time-space. Locales within geographic space, such as Kate and Sonja's homes, and the railway stations, are represented as many-sided pillars within the main column of time-space. Individuals, such as the Kate and Sonja, are represented as continuous lines, or 'life-paths' running through time space (the thick, black line, snaking its way across Figure 1). A stationary entity, for example, Kate at home, has a vertical life-path. A moving entity, such as Kate on the train, has a life-path that projects through time-space, its slope indicating speed of movement, and its horizontal deflection indicating change in position. An activity bar adjacent to the time axis indicates the activity of each individual, such as transiting and shopping, and the times at which these activities start and stop.

5 The main contribution of the chapter, then, is the performance deficits, and so the direction, goals and priories of future work. Details of possible solutions are included only to illustrate what attempting to make good a deficit may involve. To prescribe details of particular solutions, other studies are required.

6 For practical purposes, at the time of this study, GSM mobile telephones essentially provided only communication services (telephony and text messaging).

7 At the time the study was conducted, these services were voice telephony and text messaging.

8 Stress or lost opportunity was only reported for 55 percent and 40 percent of rendezvous respectively, and participants reported 5.6 rendezvous per week, giving an overall frequency of severely problematic rendezvous of around one per month.

9 In this study, about 50 percent of 31-45 year olds were married with children, compared to only 3 percent of 18-30 year olds.

10 Of course, when setting design goals, actual projects will use various sources of information, not just the performance deficits identified here. The chapter is contributing to requirements analysis, not replacing it. Similarly, the chapter is setting a direction for and prioritising design, not replacing design. These details of user interfaces in the following paragraphs make the illustration concrete, and do not derive from user performance data.

11 Position information refers to spatial coordinates, such as map reference points. Location information refers to higher level information, such as proximity to a home place, or on board a certain train.

Chapter IV
User Experience of Camera Phones in Social Contexts

Hanna Stelmaszewska
Middlesex University, UK

Bob Fields
Middlesex University, UK

Ann Blandford
University College London, UK

ABSTRACT

This chapter reports on a qualitative study into people's use of camera phones for social interaction in co-present settings. The study examined people's behaviour and positive experiences (e.g., fun, enjoyment, or excitement) when camera phones were used in different spaces (public and private). It was found that camera phones influence social practices. Three distinct practices were observed: sharing a moment now, sharing a moment later, and using photos to initiate social interaction with strangers. The knowledge obtained through the study will offer a conceptual contribution that deepens our understanding of how this emerging and evolving technology is coming to be accommodated into the leisure-related practices of its users.

INTRODUCTION

What do we know about photography? Photography has been a part of our life for a long time. We document family celebrations, important events in our lives and those of our family and friends; we take pictures when visiting museums or if we want to illustrate everyday items and people in a funny way and when we want to create stories (Mäkelä et al., 2000). It seems that photography and photos bring either smiles when reminiscing about something pleasant or tears when emotions take over. They preserve memories, capture feelings, and provide a means to communicate with others. One of the most common and enjoyable experiences is to share photos with others through story telling (Balanovic et al., 2000; Chalfen, 1987). Photos can be shared using technology and then they can be used as means for interaction with others.

Recent technological developments not only support new ways of working but also provide new mechanisms for social interaction. Mobile phones and camera phones, in particular, are examples of such technology. In the past decade, mobile phones have allowed profound changes to take place in people's behavior and practices in relation to communication (Ling, 2004), from being extensively used as a medium of verbal and text communication to one that uses pictures to facilitate people's social life. Mobile phones with integrated camera and video features have changed forever the way people communicate and interact, and have shaped both their individual and their social lives (Ito, 2005; Kato, 2005; Kindberg et al., 2005a, 2005b;Okabe, 2004; Scifo, 2004).

Although there is a vast body of literature focussing on the use of camera phones (Kindberg, et al., 2005a, 2005b; Okabe, 2004; Scifo, 2004) the issues relating to how camera phones are used to mediate social interaction between co-located users have been neglected. In this chapter, we report on the study of the collaborative use of camera phones by co-located users in various spaces.

BACKGROUND TO THE RESEARCH

In recent years, there has been substantial interest in digital photography, with a particular interest on how the digital medium facilitates sharing of images (Balanovic et al., 2000; Frohlich et al., 2002, Van House et al., 2005). Studies of sharing digital photographs include the use of Web-based systems, mobile applications, and multimedia messaging. Most of the studies focus on personal applications for sharing images remotely (Kato, 2005; Kindberg et al., 2005a; Van House et al., 2005) work on sharing images in co-present settings is in its infancy.

The issues of what people capture on mobile phones and what they do with these images were extensively investigated by Kindberg et al (2005a). They proposed a six-part taxonomy to describe the intentions behind the use of camera phone images. Intentions were grouped along two dimensions. The first intention defines whether people captured the images for affective (e.g., sentimental) or functional reasons. The second one defines social or individual intensions.

Others, such as Licoppe & Heurtin (2001) and Taylor and Harper (2003), focused on teenagers using their phones for social practices. The latter claim teenagers' practices are similar to 'gift-giving' rituals, which shape the way teenagers understand and use their mobile phones. The 'gift-giving' practices included sharing certain text messages, call-credits and even the mobile phones themselves. All these practices establish and cement allegiances and sustain rivalries (Taylor & Harper, 2003).

A field study conducted by Kato (2005) explored how the use of mobile phones/camera phones changes people's daily activities in Japan. He argues that the new ways of pervasive photo taking through camera phones allows people to document their lives on a daily basis, which can be preserved and shared as a life of a local community.

A different approach to studying mobile phone users was taken by Okabe (2004). He studied practices of Japanese camera phone users, which included personal archiving, intimate sharing, and peer-to-peer news sharing. Okabe (ibid) argues that capturing and sharing visual information cannot be understood without also understanding the social relationships and contexts within which those activities take place. Scifo (2004) provides similar views on this matter, arguing that taking photographs on camera phones and using MMS communication allows users (particularly youngsters) to identify themselves within social groups, and will intensify communication within that community.

The relevance of social relations to the uses of photographs was also identified by Van House et al. (2005). They discovered five distinct social uses of personal photos. These are: creating and maintaining social relationships, constructing personal and group memory, self-expression, and self-presentation and functional communication with self and others.

Photos could also be used for social discourse. For example, a mobile picture system (MobShare) developed by Sarvas et al. (2005) supports that by transferring photos from the phone to different devices. These include transfers (1) to another phone over the network (e.g., MMS), (2) to a PC, (3) to a network server over the network, and (4) to a printer using a cable connection or Bluetooth.

Many methods have been used to study people's uses of mobile phones, including diaries, interviews and field studies (Kato, 2005; Kindberg et al., 2005a, 2005b; Okabe, 2004; Sarvas et al., 2005). The approach employed by Sarvas et al. (2005) involved asking people to fill out a diary including all activities their performed using their camera phones. This was followed by a set of interviews focusing on photographic habits and social networking involving photography. The same methods were employed by Okabe (2004) when investigating social practices, situations and relations of the use of camera phone.

Kato (2005) applied a fieldwork study to observed and record the practices of camera phone users encouraging them not only to take pictures but also to collect and store them as visual field notes on a specially designated web site. When conducting an in-depth study of camera phone use Kindberg et al. (2005a, 2005b) applied a set of interviews asking the subject to show images that were not private from their camera phones and talk about them.

Taking inspiration from such research, semi-structured in-depth interviews and field observational studies were employed in the study reported here, which will be discussed later in this chapter.

METHODOLOGY

This study is specifically concerned with peoples' experiences when using camera phones for social interaction in a co-present setting (i.e., when participants are present at the same location at the same time). The chapter builds on an earlier more general study into peoples' experience and emotions using personal technologies such as PDAs, digital cameras and mobile phones (Stelmaszewska et al., 2005).

Because we wanted to obtain the insights of the ways people use, their camera phones as a medium for social practices we adopted Kindberg et al.'s (2005a) method of asking participants about circumstances and reasons for taking these images and their life cycle. A series of observational field studies was conducted to develop a better understanding of peoples' practices using camera phones. The use of dual methods strengthened the results obtained and provided a means of triangulation between the interviews and observations to confirm that the reported practices really did occur when the observations took place. In addition, field observations of the phenomena provided richer insights into the circumstances and contexts in which practices described in interviews actually take place.

Five students were interviewed including two PhD students, two undergraduates, and one college student, all aged between 18 and 27; all participants had been camera phone users for at least a year. Each interview took between 25 and 45 minutes and was recorded and later transcribed. The participants were asked to describe how and for what reasons they used their camera phones. The participants were also asked to show a few of the images (pictures or video) stored on their phones and encouraged to discuss where the images were taken, in what circumstances, by whom and for what reason. Also of interest was whether pictures were taken by the participant or received from another person, the means of storage and transfer employed (e.g., infrared, Bluetooth, MMS, e-mail), how long these pictures were stored, and whether they were shared with others, or retained for a private use.

The data from the field studies was gathered in a variety of public spaces, including pubs, restaurants, leisure and entertainment places, museums, and public transport (tube and buses).

The first author spent around 35 hours in public spaces observing camera phone usage. In this time, 18 individual instances of individuals and groups interacting with photos on cameras were observed and noted.

As the data gathered from interviews and field observations was of a qualitative nature, data collection and analysis was carried out iteratively. This allows for 'theoretical sampling' on the basis of concepts and themes that emerge from the analysis and allows concepts to be explored and hypotheses to be tested as they are developed from the data (Strauss & Corbin, 1998, p.46). Data from both studies was transcribed and then analyzed by first, coding it using qualitative methods to identify emerging themes, and then the themes were merged to extract the high level concepts that gave the outline of the use and practices of camera phones.

SITUATED USE OF CAMERA PHONES

The field observation study revealed many instances of people being engaged in social interaction using camera phones in different co-present settings. The in-depth interviews provided extended information to support these phenomena. The data shows the relationships between space and place as well as the photo/video sharing practices, which will be discussed in the following sections.

The concepts of place and space have been researched by many like Casey (1997), Ciolfi (2004), Dourish (2001), and Salovaara et al. (2005) just to name a few. Casey (1997) discusses this phenomena as 'space refers to abstract geometrical extension and location' whereas 'place describes our experience of being in the world and investigating a physical location or setting with meaning, memories and feelings' (cited in Ciolfi, 2004, p.1). A similar view has been taken by Dourish (2001) who gives an example of a space like a shopping street being a different kind of a place depending on the time of a day. According to Salovaara et al. (2005) the 'concepts of space and place are mutually dependent and co-occur in the context' (p. 1).

Camera Phone Use in Different Spaces

Camera phones have become a part of our lives. People carry them to work, to social events, to leisure activities, even when going shopping. Every time we use camera phones, we experience something. The experience, however, does not exist in a vacuum, but rather in a dynamic relationship with other people, places and objects (Mulder & Steen, 2005). What we experience and how camera phones are used is also determined by place and space, which will be explored in the consecutive sections.

Public Space

It appeared in the data that people use their camera phones differently depending on where they are. It was observed that when using public spaces like a tube or a bus people tend to use their camera phones for individual purposes; that includes reading and answering text messages, playing games, viewing and sorting out images, playing music or ring tones, or examining different functions on their camera phones. Interview data indicated that people do these things to overcome the feeling of boredom or simply to 'kill time' while waiting for a bus, as one of the participants (Steve) commented:

I listen to the radio … when I'm on the tube, when walking around or waiting for a bus and I don't have anything to amuse me. To amuse me, I use the calendar and the diary quite a bit. Otherwise I'd forget everyone's birthday.

Similarly, another participant (Luisa), on using camera phone on a bus, commented:

…the setting itself is boring not much inspiration to take pictures and things … you have to be with someone to do it.

It was reported in the literature that some public spaces are regulated by different means: signage, announcements and by more informal peer-base regulations (Ito, 2003, 2004; Okabe & Ito, 2005).

The former claims that these regulations are mostly exercised in public transport. Posters and signage exhort passengers from putting their feet on the seats or not smoking. The study by Okabe & Ito (2005) reported that people use email rather then voice calls when on trains and subways following 'sharing the same public space' regulations. Although, this kind of behavior was observed amongst Japanese youth population similar findings were reported by Klamer et al. (2000) who conducted a European survey investigating if the mobile phones used in public spaces disturb people.

A different kind of behavior was observed in museums (Science Museum and Natural History Museum in London). Camera phones were rarely used and only for individual purposes: receiving calls or messages, making phone calls, or texting. People treat museums as places to go on outings with friends and family, which they plan for and therefore they take a digital camera with them to capture something specific that they would like to keep as a reminder. In this case, the quality of pictures is of high importance. The comments of Maria confirm this:

…I like to take pictures of a nice scenery or … er… flowers or trees or just a really nice views or things… then I use my digital camera because of the quality of the picture.

Other public spaces like pubs, restaurants, clubs, places of entertainment and leisure provide a different social context for camera phone activities, which is in line with our previous research reported elsewhere (Stelmaszewska et al., 2005, 2006). The data illustrates that people more often engage themselves in social interaction using camera phones during gatherings with friends and family, when going out with friends or during trips or excursions with friends (see Figures 1 and 2). Most of the participants claimed that the important issue for using camera phones is to be with other people. It is people who create experiences that people enjoy, as Adam noted:

When you have other people around you then you have a different kind of experience.… you are more likely to do silly things. So then you take pictures and when you view them you can laugh and have fun. When you are on your own … no, you don't do these things. You need to have people around you to have fun.

Private Space

A similar behavior was reported when groups of participants use their camera phones in private spaces (e.g., homes or cars); that is people took pictures or videos of friends, members of family or even themselves behaving funny or silly and

Figure 1. A girl sitting with her family and taking picture of the artist playing

Figure 2. People taking photos of the pantomime artist

then shared them with others co-present or they viewed pictures and videos taken previously. The comment from Adam supports this view:

...so what we did was just running through clips and passing them from one group of people to another … [laughing] this was funny… I like to take pictures of funny situations and when my friends are drunk they do funny things so we go back and try to remember what happen and we always have a good laugh. Sometimes we like to compare who managed to take the most funny shots… it is really funny seeing people doing crazy things.

Since the camera screens are small and do not support easy and clear viewing for a group of people when sharing pictures in the home environment, people often made use of external display technology, such as TV or computer. This issue will be explored further in the next section, *'Sharing a moment later.'*

As discussed in this section peoples' use of camera phones changes in relation to the space they are in; private vs. public. It was found that people's practices when using camera phones differs in different spaces. The next section will discuss this phenomenon in more detail.

Social Uses of Camera Phones

Camera phones have been used for individual as well as group purposes. Consistent with other studies (Kindberg et al., 2005a, 2005b) we found that people take photos for individual purposes that include creating memories and evocations of special events, trips, holidays, or beautiful landscapes. A common practice is to share images with friends and family, in a way that is deeply embedded in social interaction (Stelmaszewska et al., 2005, 2006). Sharing digital photos is often done remotely via email or by posting them on the web (Counts & Fellheimer, 2004; Stelmaszewska et al., 2005). Despite the growing popularity of using web-based applications and services (e.g., Flickr, YouTube, or Mobido) that allow their users to share photos there were no accounts reported using these services by the participants involved in this study.

However, we observed other practices that occur in co-present social contexts. These include 'sharing a moment now,' 'sharing a moment later,' or using photos to initiate social interaction with strangers.

'Sharing a Moment Now'

This study shows a different way people share photos taken on a camera phone that appears to be less about evoking or recreating an event or scene after the fact, and more about augmenting that event as it happens. It was observed that people take a 'spur of the moment' photo or video and share it with people who are present at the same location at the same time. People reported having fun when taking photos or videos of their friends behaving funnily and then viewing them collectively at the location. This kind of behavior seems to motivate and shape social interaction, as Adam reported:

…she was happy and funny (referring to a friend) … far too engaged with dancing to notice what was happening around her … and I just thought that I'll just take that picture. … there were few of us friends so then I showed them and then other friends were taking more pictures of her dancing and we were waiting for her to realize what was going on… we were all taking pictures of her… we shared all the pictures and picked out the funniest ones. It was so funny because she couldn't believe that we did that and she didn't even notice it.

Whereas Lucy said:

When I'm out with my friends then I'll definitely use it (referring to a camera phone). … Sometimes I take pictures of my friends and then we'll sit down and go through them selecting the best once.

Data shows that photos were used for functional purposes as well, which is consistent with the findings of other research (e.g., Kindberg et al., 2005a; Van House et al., 2005). It was observed that when on a trip, people took a picture of a map displayed by a leader and then pursued his instructions using a display on their camera phones. This kind of activity allowed every person within the group to see clearly the map and use it for further reference.

Another common practice observed and reported by participants was to transfer photos between phones using the Bluetooth technology so that everybody concerned could store and use them when needed. The following observed episode is a typical example:

Episode 1: Pub, evening

Ten people are sitting at the table (three females and 7 males). Jim takes the camera phone out of his pocket and plays with it.

Jim: 'I have something really cool to show you.' He does something with his phone. After a while Jim said: 'OK, I've got it.' He plays the video and passes his phone over to a neighbor, Roy.

Jim: 'Just press the button.' Roy plays the video and moves the phone towards another male, Paul. Another male, Martin moves from his seat and stands behind Roy and Paul watching the video clip.

Martin: 'I want this clip. Can you Bluetooth it?'

Jim: 'Yeah' Jim takes his phone back from Roy and sets up the Bluetooth. Martin does the same on his phone. After a short while Jim transfers the clip over to Martin's phone.

However, it appeared that some people found it difficult to use it and either abandon the transfer or asked for help. When discussing issues related to managing pictures on the phone Maria said:

I Bluetooth them … I can do it now but I had to ask my friend to show me how to do it so I'm OK now.

'Sharing a Moment Later'

When people who you want to share photos with are around, it creates opportunities for social interaction to take place so that people can enjoy the moment of sharing pictures together. What happens when they are not around? Other studies reported this kind of practice; that is to view the photos when the occasion arises, and not immediately after they have been taken. For example, Okabe (2004) described situations where people show their friends the photos from their archives (photo gallery) on occasions that they get together.

A co-present social interaction was reported to be associated with participants' experience when viewing pictures or videos stored on individual's phones but taken previously (not at the time of gathering). The intentions behind it were reported to include sharing memories of special events, reporting on events to those who were absent at the time of events, or creating and sharing a documentary of a friendship or family life as Maria remarked:

with the cam_phone I can capture the moment … and being able to view them later will bring all the memories and the fact that those pictures can be shared … so people can have fun.

People were more inclined to use photos for storytelling, which is in line with (Balanovic et al., 2000; Kindberg et al., 2005a) and, as suggested by Fox (2001) and Vincent & Harper (2003), mobile phones have been used to maintain personal relationships between friends and family. Since camera phones are becoming a part of our everyday lives, it is not surprising that the same behavior was observed in the context of camera phone use when photos or videos were shared during social gatherings.

However, given that phone screens were claimed to be very small it was common amongst participants to use other media like computer or

TV to display photos in order to improve their visibility and enhance the experience of people participating. Adam reported:

I transferred them onto my computer … I'm quite organized with my pictures so I categorize them and put them in kind of albums and sometimes when I'm with friends we like to go through pictures and have fun.

Maria commented:

…sometimes what we do is we Bluetooth to transfer our pictures to one of our computers and then have a slide show so everybody can see it …you see the phone screens are very small and if we all want to have fun we need to see those pictures simultaneously. With camera phones we can't see it clearly if there are more then two or three people looking. It's just not enough space …

Sharing photos at co-present settings proved to be a way of social interaction that brings fun and joy to people's lives. The remarks of an interviewee, Steven, appear to confirm this point:

I'll show them (referring to family) what I managed to capture and then we have a good laugh.

Supporting the view, Lucy commented:

…you take pictures and when you view them you can laugh and have fun.

Ito & Okabe (2003, p.6) claim that: "Mobile phones … define new technosocial situations and new boundaries of identity and place … create new kinds of bounded places." We argue that camera phones go beyond that. When people view pictures together and tell the story behind them, they are transported to the place and space where those pictures were taken. Pictures conjure memories, feelings, and emotions and evoke sensations associated with the events that were photographed. Lee, another study participant, remarked when showing pictures from a group trip:

… The first dive was really s…. it was sooo cold, remember, … and we didn't see much... The vis was absolutely s…. yeah and then we had to get warmer ha, ha, ha …

Comments from other participants suggest the same:

Adam: *… when you are having a good time you don't always know what's happening around you. … I don't always know what everybody is doing so I miss a lot of stuff but when we view all the pictures taken during a particular party or we go for a short trip together … so only then you really can see what happened. We really like doing that.*

Maria: *… you can not only see the pictures but there are always some stories behind every picture. … so later when you show the pictures everybody gets involve and just add a story to it and that's great. I like it. And others who were not there can feel like they were there err...kind of.*

Social Interaction with Strangers

Studies reported by Weilenmann and Larson (2002) explored the collaborative nature of mobile phones use in local social interaction amongst teenagers. They suggest that mobile phones are often shared in different forms including: minimal form of sharing (SMS messages), taking turns (several people handling a phone), borrowing and lending of phones, and sharing with unknown others. The latter involves the phones being handled by teenagers who are unacquainted until one of them makes the initial contact. Weilenamm and Larson (2002) describe practices of teenagers (boys giving girls their mobile phone) to enter their phone numbers. This kind of social interaction is similar to the one that emerged from our studies.

Social interaction can coalesce around different media, from text and graphics, to interactive games (Stelmaszewska et al., 2005, 2006). Such interactions often occur between friends or family members sharing the same technology (i.e., computer, digital camera or mobile/camera phone).

However, a striking finding was that camera phones were used as a new channel and medium for initiating social interaction with strangers. It was reported that people take photos of others (whom they like) in order to show their interest, introduce themselves, or simply start a new social relationship.

The comment from Luisa supports this claim:

I was at the Harvester, a restaurant/pub thing, …and there was a small window with glass between it looking like a fake door and the guys were looking through that doing (mimicking facial expressions) and then I saw one holding his camera phone against one of the window things and there was a picture of me going (shows facial expression) and I didn't know that they were taking it … I didn't really mind. It's a good humor… it was kind of friendly, sort of vague flirting without talking … just taking pictures.

So does another comment by Maria:

We were in the bar … having fun and there was this guy dancing [laughing] kind of a very funny dance … almost like an American Indian kind of dance … and one of the girls from our group took a photo of him because she liked him and she was showing it to us so instead of looking at him we could see his picture … and when he saw her taking pictures of him he did the same to her… the whole situation was funny … at least we had fun watching them two taking pictures of each other instead of talking …

This kind of behavior typically occurred in public spaces such as pubs, bars, or clubs where people usually gather for social events, and interaction with others is a part of the entertainment. In our study, the focus was on social interaction that took place through and around digital photos. Such interaction is not always appreciated by those involved. Some participants felt offended and annoyed with those taking photos without obtaining agreement. For example, Lucy noted:

I don't know if I would be offended so much. I think it depends what for … sometimes you get photographers going like around pubs and clubs… and I never said yes to the photo. The other night when I was there with my friend and this group of guys we met before errr … this guy said: 'Oh yeah, let's get a picture' but we went like: 'no, we really don't want to.' And they had one done anyway and this kind of annoyed me a bit because … it's fair they wanted the picture of us but we didn't really want to be in it. … I think it depends how much choice you are given as whether or not you want your photo taken.

It appeared that pictures are not the only phone-related way people try to 'chat up' others. Phone features like Bluetooth can be used to connect to strangers and initiate communication. This kind of behavior was observed in public places (pubs, restaurants, bars). The practice was to switch on the Bluetooth and ask others (whoever is picked up by the Bluetooth) to activate the connection. However, this kind of interaction often raised some suspicions, as people did not know who wants to 'chat up' to them. Here is an extract from one of the participants expressing his concerns:

… someone wants me to activate the connection … but what do I do … I don't want any 'Boss' [the name of the Bluetooth connection] connecting to my phone. What if they do something to my phone?

The fact that people do not see the 'talker' and they do not have the full control of who they interact with seems to be a barrier to engaging in interaction with a stranger.

It seems that communication takes place not only through technology but also alongside it, a finding that is consistent with our earlier studies (Stelmaszewska et al., 2005). Moreover, Van House et al. (2005) argue that technology (e.g., online photo blogs) is used to create new social relationships. Although this study is at an early stage and further evidence is required, we suggest that camera phones provide new channels and foci for social interaction within co-present settings.

Barriers to Sharing

Although camera phones appear to be a new medium for social interaction that is enjoyable and fun, they are not without problems that limit the extent to which they are used. The data illustrates that people experience different kinds of trouble that hinder their experience or make it impossible for sharing to happen.

Firstly, the lack of compatibility between different camera phones stops people from sending photos. Several participants reported not using MMS features because it was difficult to use. In addition, people often know (not always) that those who they want to send pictures to will not be able to retrieve them as was commented by Luisa:

… none of mine friends really do this … you have to have the same phone or something to be able to send it and for them not to just say: 'message not being able to deliver or whatever.' Some people tried to send pictures on my phone but I never got them.

Secondly, for many camera phone users it is difficult to send pictures either via MMS or Bluetooth. People reported having difficulties to find the functions to do so or they could not set them up (in case of the Bluetooth—see comments in the section on 'Sharing a moment now').

Another barrier to sharing photos was the lack of a quick and easy way to find archived pictures. People spent time, sometimes a long time, trying to find the pictures they wanted to share with their friends. This caused frustration and dissatisfaction as Jim said:

Where is it?!!! S… Hrrrrrrrrrr

Quick access to camera functionality and photo image features is an important issue in a context of sharing and it raised concerns amongst participants as Maria noted:

… one of my friends helped me to set it up so I can use it by pressing just a couple of buttons instead of going through menus and stuff. It was horrible. I missed so many great pictures because of that and I was very upset about it. … it's very important. I could have so many great pictures but couldn't find the camera function on my phone ... it was very frustrating.

All these barriers affect not only experience of camera phone users but also their engagement in social interaction. So providing functionality that is transparent and supports users sharing activities is of a paramount importance when designing systems. It might also enhance the use of camera phones by creating pleasurable and fun experiences instead of satisfying only functional purposes.

DISCUSSION AND CONCLUSION

It seems that phone technology is moving from facilitating its original primary goal, supporting distance communication, to supporting new ways of social interaction that happens through sharing activities (photos and videos) as well as providing bridges between contexts. When people share photos or videos, they are transported from the context of a present space (pub, restaurant, or home) to the one that a specific photo or video clip conjures up.

In addition to providing resources for communication and interaction, camera phones have been used as a kind of archive of a personal life, a viewpoint on the world, or a collection of fragments and stories of everyday life. Okabe (2004) suggests that photos are often taken for purely personal consumption, whereas text messages are generally created with the intent to share with others. However, the findings from this study contradict Okabe's claim; people often take photos with the intention to share them with others, which is a more selective and intimate activity than sharing text.

When technologies are used in different places and spaces they become part of a specific environment and this often shapes the use of technology and experiences connected to it. As a consequence of this, technologies are often used in unexpected ways (Taylor & Harper, 2003). In the case of this study, these ways are 'sharing the moment now,' 'sharing the moment later' and using camera phones for 'social interaction with strangers.'

This chapter has described distinctive practices of camera phone users occurring in co-present settings, and how these practices change in relation to the place and space in which they were used. It has been argued that camera phones provide a new medium through which people can sustain and enrich they social interaction through taking and sharing photo images or videos. However, these activities are inseparable from social relations and context, which is in line with Okabe's (2004) and Scifo's (2004) findings. Moreover, we argue that this study provides a better understanding of how this emerging and evolving technology facilitates social interaction in the leisure-related practices of its users.

We agree with Rettie's (2005) view that mobile phone communication affects the role of space and we have shown that camera phones go beyond this: they bring people together, creating experiences through social interaction. No other technology has supported this to such an extent, and to so many people. The multi-functionality of camera phones provides a different means of social interaction, which is unique to a place and space.

More generally, when designing camera phones that facilitate social interaction, understanding of emerging uses, practices and social activities is essential for the effective design of camera phones and related systems. Moreover, identifying problems within existing systems might be a good starting point for discussing user requirements, helping designers to develop systems that fulfill utilitarian as well as user experience needs.

Although the notions of 'sharing' might be a new phenomenon it is a manifestation and reflection of needs that relate to social identity (Scifo, 2004; Taylor & Harper, 2002) and are shaped by social context (Okabe, 2004; Stelmaszewska et al., 2005, 2006). This study is part of an ongoing effort to explore issues related to the use of camera phones for social interaction within co-present settings, and further studies will be required to investigate what affects such interaction, how camera phones' design, usability and context of use influence the nature of users' experience.

Furthermore, more work is needed to identify and understand problems when camera phones are used for social interaction, and how we can improve the design of camera phones so that they can evoke experiences such as pleasure, excitement, or fun.

ACKNOWLEDGMENT

We would like to thank all anonymous participants who took part in this study.

REFERENCES

Balanovic, M., Chu, L., & Wolff, G. J. (2000). Storytelling with digital photographs. In *Proceedings of CHI 2000: Conference on Human Factors in Computing Systems* (pp. 564-571). ACM Press.

Casey, E. S. (1997). *The fate of place: A philosophical history*. Berkeley: University of California Press.

Chalfen, R. (1987). *Snapshot version of life*. Bowling Green, OH: Bowling Green State University Press.

Ciolfi, L. (2004). *Digitally making places: An observational study of people's experiences of an interactive museum exhibition*. Paper presented at the Proceedings of the 2nd workshop on 'Space, Satiality and Technologies.' Edinburgh, UK.

Counts, S., & Fellheimer, E. (2004, April 24-29). Supporting social presence through lightweight photo sharing on and off the desktop. In *Proceedings of CHI 2004: Conference on Human Factors in Computing Systems* (pp. 599-606). Vienna: ACM Press.

Dourish, P. (2001). *Where the action is: The foundation of embodied interaction*. Cambridge, MA: MIT-Press.

Fox, K. (2001). Evolution, alienation and gossip: The role of mobile telecommunications in 21st century. *Social Issues Research Centre Report.* Retrieved from http://www.sirc.org

Frohlich, D., Kuchinsky, A., Pering, C., Don, A., & Ariss, S. (2002). Requirements for photoware. *Proceedings of the 2002 ACM conference on Computer Supported Cooperative Work* (pp. 166-175). ACM Press.

Ito, M. (2003). *A new set of social rules for a newly wireless society. Japan Media Review.* Retrieved July 28, 2006 from http://www.ojr.org/japan/wireless/1043770650.php

Ito, M. (2004). *Personal Portable Pedestrian: Lesson from Japanese Mobile Phone Use*. Paper presented at The 24 International Conference on Mobile Communication Social Change. Seoul, Korea.

Ito, M., & Okabe, D. (2003, June 22-24). Mobile phones, Japanese youth, and the re-placement of social contact. In R. Ling (Ed.), *Front stage—back stage: Mobile communication and the renegotiation of the public sphere.* Grimstad, Norway.

Ito, M., & Okabe, D. (2005). Technosocial situations: Emergent structurings of mobile email use. In M. Ito, Okabe, D. & Matsuda, M. (Ed.), *Personal, portable, pedestrian: Mobile phones in japanese life*.

Kato, F. (2005, April 28-30). *Seeing the "seeing" of others: Conducting a field study with mobile phones/mobile cameras.* Paper presented at the T-Mobile conference, Budapest, Hungary.

Kindberg, T., Spasojevic, M., Fleck, R., & Sellen, A. (2005a). An in-depth study of camera phone use. *Pervasive Computing, 4*(2),42-50.

Kindberg, T., Spasojevic, M., Fleck, R., & Sellen, A. (2005b). I saw this and thought of you: Some social uses of camera phones. *In CHI '05 extended abstracts on Human factors in computing systems* (pp. 1545-1548). ACM Press.

Klamer, L., Haddon, L., & Ling, R. (2000). The qualitative analysis of ICTs and mobility, time stress and social networking. (No. P-903): EURESCOM.

Licoppe, C., & Heurtin, J.P. (2001). Managing one's availability to telephone communication through mobile phones: A French case study of the development dynamics of mobile phone use. *Personal and Ubiquitous Computing, 5*(2), 99-108.

Ling, R. (2004). *The mobile connection: The cell phone's impact on society.* Morgan Kaufmann.

Mäkelä, A., Giller, V., Tscheligi, M., & Sefelin, R. (2000). Joking, storytelling, artsharing, expressing affection: A field trial of how children and their social network communicate with digital images in leisure time. In *Proceedings of CHI 2000: Conference on Human Factors in Computing Systems* (pp. 548-555). ACM Press.

Mulder, I., & Steen, M. (2005, May 11). *Mixed emotions, mixed methods: Conceptualising experience of we-centric context-aware adaptive mobile services.* Paper presented at the Pervasive 2005. Presented at workshop on User Experience Design for Pervasive Computing, Munich, Germany.

Okabe, D. (2004, October 18-19). *Emergent Social practices, situations and relations through everyday camera phone use*. Paper presented at the International Conference on Mobile Communication, Seoul, Korea.

Okabe, D., & Ito, M. (2005). Ketai and public transportation. In M. Ito, Okabe, D., & Matsuda, M. (Eds.), *Personal, Portable, Pedestrian: Mobile Phones in Japanese Life*. Cambridge: MIT.

Rettie, R. M. (2005). Presence and embodiment in mobile phone communication. *PsychNology Journal*, *3*(1), 16-34.

Salovaara, A., Kurvinen, E., & Jacucci, G. (2005, September 12-16). *On space and place in mobile settings.* Paper presented at the Interact '05, Rome.

Sarvas, R., Oulasvirta, A., & Jacucci, G. (2005). Building social discourse around mobile photos—A systematic perspective. *Proceedings of the 7th international conference on Human computer interaction with mobile devices & services* (pp. 31-38). ACM Press.

Scifo, B. (2004, June 10-12). *The domestication of camera-phone and MMS communication: The early experience of young Italians.* Paper presented at the T-Mobile Conference, Hungary.

Stelmaszewska, H., Fields. B., & Blandford, A. (2005, September 5-9). *Emotion and technology: An empirical study.* Paper presented at the Emotions in HCI design workshop at HCI '05.

Stelmaszewska, H., Fields, B., & Blandford, A. (2006, September 11-15). Camera phone use in social context. In *proceedings of HCI 2006* (Vol. 2, pp. 88-92), Queen Mary, University of London, UK.

Strauss, A., & Corbin, J. (1998). *Basics of qualitative research. Techniques and procedures for developing grounded theory.* Newbury Park: Sage.

Taylor, A., & Harper, R. (2002, April 20-25). *Age-old practices in the 'new world': A study of gift-giving between teenage mobile phone users.* Paper presented at the CHI, Minneapolis, MN, USA.

Taylor, A., & Harper, R. (2003). The gift of the grab: A design oriented sociology of young people's use of mobiles. *Journal of Computer-Supported Cooperative Work, 12*(3), 267-296.

Van House, N., Davis, M., Ames, M., Finn, M., & Viswanathan, V. (2005, 2-7 April). The uses of personal networked digital imaging: An empirical study of cameraphone photos and sharing. *CHI '05 extended abstracts on human factors in computing systems* (pp.1853-1856). ACM Press.

Vincent, J., & Harper, R. (2003). *Social Shaping of UMTS: Preparing the 3G Customer.* University of Surrey, Digital World Research Centre. Retrieved July 26, 2006, from http://www.dwrc.surrey.ac.uk

Weilenamann, A., Larsson, C. (2002). Local use and sharing of mobile phones. In B. Brown, Green, N., and Harper, R. (Ed.), *Wireless world: Social and interactional aspects of the mobile age* (pp. 92-107): Springer.

KEY TERMS

Bluetooth: A wireless protocol that is used to connect compliant devices that are in close proximity with each other in order to transfer information between them. Bluetooth is commonly used with phones, hand-held computing devices, laptops, PCs, printers, digital cameras.

Camera Phone: A mobile phone with a camera built-in that allows the user to take pictures and share them instantly and automatically via integrated infrastructure provided by the network carrier. Camera phones can transfer pictures via Bluetooth, Infrared, or MMS messaging system.

Co-Present Interaction: Interaction that happens between two or more people that are physically present at the same time and location.

Digital Photo Sharing: An activity of two or more people, who share images by showing pictures to others. Sharing digital photos can occur at the co-present location or remotely. The former happens using different devices like camera phone screen, digital cameras, TV screen, or computer screen. The latter is often done via email or by posting them on the web.

Digital Photography: A type of photography where pictures are taken on digital cameras or camera phones. Images can be viewed, edited, stored, or shared with others using different means of communication medium such as email, Web-based applications and services, Bluetooth, Infra-red, MMS, computers or TV screens.

Field Observation Studies: A qualitative data collection method, which is used to observed naturally occurring behavior of people in their natural settings. The data can be gathered in a form of: film or video recording, still camera, audio type (to record spoken observation), or hand-written note taking.

Qualitative Data Analysis: A collection of methods for analyzing qualitative data, such as interviews or field notes. One example of such method is Grounded Theory, which is used to generate theory through the data gathering and analysis. Data is sorted to produce categories and themes of concepts emerging from the data.

Social Interaction: Interaction that happens between individuals typically mediated by, or in the presence of technological artifacts.

Theoretical Sampling: The process of data collection for generating theory where the researcher collects, codes and analyses data and makes decisions about what data to collect next. Researchers consciously select additional cases to be studied according to the potential for developing new insights or expanding and refining those already gained. Sampling decisions depend on analysis of data obtained, which relate to the developing theory.

Triangulation: The application and combination of at least two research methods or data gathering exercises to research the same phenomena in order to cross-checking one result against another, and increasing the reliability of the results.

User Experience: A term that is used to describe the overall experience and satisfaction of a user while using a product or system.

Chapter V
Interaction Design for Personal Photo Management on a Mobile Device

Hyowon Lee
Dublin City University, Ireland

Cathal Gurrin
Dublin City University, Ireland

Gareth J.F. Jones
Dublin City University, Ireland

Alan F. Smeaton
Dublin City University, Ireland

ABSTRACT

This chapter explores some of the technological elements that will greatly enhance user interaction with personal photos on mobile devices in the near future. It reviews major technological innovations that have taken place in recent years which are contributing to re-shaping people's personal photo management behavior and thus their needs, and presents an overview of the major design issues in supporting these for mobile access. It then introduces the currently very active research area of content-based image analysis and context-awareness. These technologies are becoming an important factor in improving mobile interaction by assisting automatic annotation and organization of photos, thus reducing the chore of manual input on mobile devices. Considering the pace of the rapid increases in the number of digital photos stored on our digital cameras, camera phones and online photoware sites, the authors believe that the subsequent benefits from this line of research will become a crucial factor in helping to design efficient and satisfying mobile interfaces for personal photo management systems.

INTRODUCTION

Long before digital technology came into everyday use, people have been managing personal photos with varying degrees of effort. Individuals' photo management strategies ranged from stacking photos in shoe boxes to carefully placing them into a series of photo albums with detailed notes of where and when each photo was taken or a witty caption beside it. Reminiscing and story-telling past events that have been visually recorded in personal photos is a highly-valued activity for many people. This gives meaning to the person's past events and also works as a socially-binding and relationship-enhancing device at gatherings of family or friends. With the Internet revolution, and the arrival of inexpensive digital cameras, people's photo organizing and sharing behavior has been evolving as new technologies allow different ways of managing photo collections. This is exemplified with online photoware applications such as Flickr[1], with which people can now upload personal photos taken from their digital cameras onto a shared web space on which collaborative annotation, browsing and sharing photos with other people is possible.

Another aspect of the development of digital photography is that people's behavior in capture of photos is changing as well. In particular, due to the low cost and ease of capture nowadays people are taking many more photos than in the past. This is possibly best illustrated by the ubiquity of camera phones, mobile devices that can be used as digital camera as well as a phone. Many people carry their phone with them at all times meaning that they can capture their everyday lives and holiday scenes whenever they want. This change in capture behavior can also have a significant impact on people's personal photo management activity. Once captured, the phone can be used to send photos to a friend's mobile phone or to upload them to a public Website for instant sharing and receiving comments back. This means that when designing personal photo management tools, we should consider the implications of the changes in user photo capture behavior arising from the emergence of the ubiquitous availability of the means of photo capture. For example, there is a need to design specific user-interfaces for photo management on a camera phone itself. A camera phone may be used merely as a capture device that takes photos and stores them, to be copied later to a PC for further photo management. However, the quality of screens now commonly available on mobile phones means that it is quite reasonable to look to design tools that enable users to organize, annotate and browse photos *on* the mobile phone itself. Between these two extreme cases, there is a spectrum of varying degrees to which a camera phone or other mobile device can be integrated into overall photo management functions and tasks, effectively a continuum of trade-off among technological resources and the user's effort and time. For example, due to the difficulty of text input on a camera phone arising from physical constraints, it may be easier for the user to fully annotate photos after copying them on to a desktop PC at home. Even so, a user in some situations might still want to make the effort to annotate their photos using the mobile and to send them to a friend for the benefit of its immediacy and not having to do the extra work of copying photos to a PC at home in the evening before performing the annotation. On occasion some users will want to bulk-upload a large number of photos taken at a party directly to a website without any annotation, e.g., to share with close friends. Depending on the design decisions on the allocation of photo management tasks for different devices, the optimal user-interface for such tasks on the mobile device will vary. Currently available interfaces on camera phones and digital cameras for photo management illustrate this possible diversity of user task requirements.

Particular challenges faced in designing and evaluating mobile interfaces for personal photo management arise due to, among other things, the following:

- New technology regularly emerges and applications constantly evolve.
- Mobile users are difficult to observe.

Consequently, it is difficult to rely on the traditional system development cycle of user study, user needs, and requirements establishment, followed by prototyping and evaluation. By the time this established process has been completed for an application, a new technological innovation may have appeared bringing in a new possible line of products to be developed. Conducting a user study on mobile devices can pose significant problems compared to desktop-based systems because a proper observation of users using a mobile device is technically difficult. When the user is on the move, we are simply unable to hook up an observation camera and recorder, and keep following the user while he or she is out and about using the device.

However, recently study methodologies have been developing to cope with these difficulties. For example, quicker and cheaper prototyping techniques are starting to be adopted to keep pace with changing technology (e.g., PC-based simulated prototypes or using a general mobile platform (Jones & Marsden, 2006, p. 179)). In another example, user studies are emerging which adopt light ethnographic and indirect observations using diaries and self-reporting to cope with test users on the move (Palen & Salzman, 2002; Pascoe et al., 2000; Perry et al., 2001). More specifically on the use of *camera phones*, some early longitudinal user studies have appeared (Kindberg et al., 2005; Sarvas et al., 2005; Van House et al., 2005) that aim to better understand the different motivations and current practices in camera phone use. These studies use diaries and interviews to capture usage data from 30-60 users within a period ranging from one to two months.

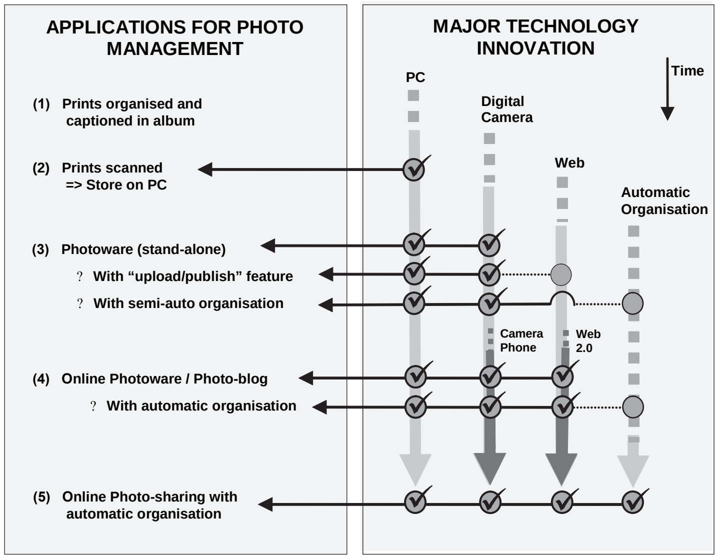
In understanding the current status of mobile interaction design for personal photo management and in setting the right direction for future applications, we need to look at the way such applications have been developed so far. As will be described in more detail in the following sections, applications for personal photo management have been incrementally shaped by the major technology innovations that have appeared during the last two decades or so. By looking at other new technologies, which are likely to be available in the near future, we can roughly determine what kind of personal photo applications or requirements will emerge in which mobile devices are an important component. The primary technology that we anticipate will appear is the *automatic organization* of the photos. The need for and adoption of this technology will be driven by the very rapid increase in the number of digital photos that users will accumulate on camera phones, desktop PCs, and online photoware. The sheer volume of photos means that it is increasingly difficult and frankly becomes unrealistic to manually annotate these photos.

Fortunately, by leveraging context data such as the time and location of photo capture, the bulk of the organization task can be done automatically. In addition, content-based image analysis techniques, although considered still not mature enough for many other applications, are also proving a very promising element that can further contribute to effective automatic organization and annotation of large photo collections. Use of these organization automation tools reduces the user's annotation burden to such an extent that a user working with large digital photo collections can focus on enjoyable browsing, searching and sharing tasks, rather than the nuisance of file extension and ongoing manual annotation. In this chapter we explore how mobile interaction design for personal photo management can take advantage of these emerging technological factors to overcome potential interaction design problems for photo management applications on a mobile device.

BACKGROUND: TECHNOLOGY TREND AND PHOTO MANAGEMENT ON MOBILE

Starting from physically printed photos, which are manually organized in an album, moving to current camera phones to capture, annotate, and browse digital pictures, the way we manage personal photos is evolving very rapidly. Current emerging applications have largely been geared to the major technological innovations that have occurred. These include the wide uptake of PCs

Figure 1. Supporting personal photo management at different stages of technological development



to the general public since the mid-80s, the World Wide Web and inexpensive digital cameras that have become commonplace since the mid- and late-90s. Each of these has brought a set of new ways of managing personal photos and the applications (with their associated features) have been quickly developed and used. This is illustrated in Figure 1.

On the right side of the figure, four vertical arrows indicate the major technologies: PCs, digital cameras and camera phones, Web and Web 2.0 technology, and finally a newly appearing technology, Automatic Organization. The first three of these have already shaped the way people manage their personal photos, in effect the *enabling technology* for personal photo management. These technologies were not mature when first emerging (thus the dotted line at the start of each of the vertical lines), but in time have become reliable enough to be taken up as full applications. The rows on the left side of the diagram show the way people manage their photos taking advantage of these technologies. For each row the cross point with the vertical lines indicates which technology has been used (circle with a tick mark). Circles without a tick mark indicate cases where the uptake of the technology has only been partial or done in a work-around way due to insufficient development of this particular technology, or of our understanding of its potential at that time. The last row is the envisaged future photo applications that take full advantage of the four technologies including Automatic Organization.

In this section we briefly go through each of the rows as numbered in the figure, highlighting how the enabling technologies influenced its development, what specific features became possible due to such technologies, and where mobile device and associated user-interface design issues arise from this.

Prints Organized and Captioned in an Album

Since the camera became a common household gadget, the most traditional way to manage personal photos was to get film developed and then select well-taken prints to insert into an album. Because the user selected only good photos for inclusion in a photo album, it was most likely to contain high quality or useful photos (Rodden, 1999). The major feature of an album is the grouping or organization into pages and the *captioning*, often appearing beside the chosen photos where the user adds a short description of the photo, often with a humorous comment. When people meet at a family event or party, they might gather around the album and flip through pages discussing and appreciating photos and their captions; good photos were also ordered for reprints to be given or posted to families and friends for sharing. These social activities have been popular as the main ways of browsing and sharing personal photos (Frohlich et al., 2002) and have often been adapted in photoware as a metaphor.

Prints Scanned and Stored on a PC

Before digital cameras became inexpensive and commonplace, those who owned PCs could scan prints of physical photos to store them digitally on their PCs. Preservation, digital enhancement or novelty were more likely motives rather than as a true replacement of chemical photos. Sharing was possible by copying files onto a disk and passing it to somebody who also owned a PC or more recently sending the photos as an email attachment (Frohlich et al., 2002; Rodden & Wood, 2003), but the PC did not come to be used as a tool for photo management until people started taking photos with digital cameras.

Stand-Alone Photoware

The availability of inexpensive digital cameras meant people could now capture their interesting events directly into digital format and copy them onto their PC. As the cost of capture is virtually zero and unwanted photos are easily deleted, the average number of photos taken increased dramatically. However, digital photos stored in PC directories often have cryptic file names such as "1430XX23-02.jpg" which have been generated automatically by the camera. The large quantity of such photos means that users often do not attempt to make the effort to rename them with more meaningful titles. It is thus not possible for users to find individual photos based on filenames.

As the Web started to become more common, some users manually created Web pages containing their photos to share with family and friends. However, this required much time and effort to find and select photos from within their collections and then to generate the Web pages, and only carefully selected photos were made available for sharing by users interested in both web and digital photography technologies.

This situation improved with the introduction of *photoware* software that imports photos from a digital camera and supports easy management of photos by allowing the grouping, sorting by date, annotation, and subsequently allowing searching and browsing. As people started accumulating large numbers of photos the utility of photoware grew. Examples of popular photoware include Photoshop Album[2], ACDSee[3] and Picasa[4].

Some experimental photoware systems support automatic grouping of photos based on time/date as the photos are imported from the camera (O'Hare et al., 2006), removing some of the organizing burden from the user while at the same time motivating higher quality annotation (Kustanowitz & Shneiderman, 2004). Other systems feature a convenient "upload" or "publish" feature whereby after organizing photos on a PC, the system generates attractive Web pages with the selected photos and captions.

Currently a large number of similar photoware tools are available each providing some combination of features including photo manipulation (rotation, brightness/contrast change, sharpening, red-eye removal, adding visual effects, etc.), organization (adding titles and descriptions, fast photo tagging, searching for duplicate photos, synchronising directories, etc.) and browsing and searching (thumbnail and full-screen views, slideshows, easy zoom-in/out, search by annotation, view colour histogram, etc.).

Online Photoware for Sharing and Photo-Blogging

With the advent of Web 2.0 in which the web is itself a platform, a highly interactive user-interface can be realized directly within the Web browser. Conventional stand-alone software applications such as word processors, time schedulers and e-mail clients are now available online, and photoware applications can be deployed in this way as well. Organization, annotation and other useful photo manipulation such as rotation and cropping, as well as various photo collection visualizations are now featured in online photoware in which the user directly interacts with a Web-based interface.

In addition, the proliferation of *camera phones* has resulted in ubiquitous use of capturing and sharing of digital photos, spurring ever higher quantities of digital photos to be taken. In parallel, some mobile phone services have started allowing digital photos taken on a camera phone to be directly uploaded to online photoware applications and shared with other users, bypassing the step of copying photos onto a PC. Use of a camera phone for saving, organizing, annotating, browsing, searching, and sharing photos raises multitudes of user interaction issues, as we will see in more detail in the next section.

Finally, some users have started regular online posting of their daily photos with annotations allowing a community of online citizens to comment on their photos and annotations, referred to as *photo-blogging*, merging the uploading of digital photos with the features of text blogging.

Emerging from the combination of these digital and online technologies, we have popular online photoware such as Flickr[5] and Yahoo! 360[6] providing highly interactive photoware features, while focusing on the online community and sharing of photos with such features as searching over community photos, popular photos of the day, commenting and photo grouping across users. Although, not yet a mature genre, the early design guidelines for online photoware can already be found drawn from general design principles and developers' experiences (Frohlich et al., 2002).

Online Photo-Sharing with Automatic Organization

Manually organizing photos into groups and subgroups and annotating them one by one is at times pleasant, but as our photo-taking habits change from being selective to taking as many photos as possible, creating meaningful captions for individual photos becomes more and more time consuming. However, to be able to subsequently access the photos, appropriate organization and annotation of each photo is crucial. While the subjective nature of indexing a visual medium is a problem and in itself an important research area (Enser, 1995), to help a user index hundreds or thousands of personal photos for efficient searching and browsing is a challenging problem both in terms of design and technique.

Photoware which *automatically* organizes and annotates photos for the user is an attractive possibility, and as with any other digital library project where a system automatically indexes the documents in the database, is becoming more and more feasible with technical advancements that are happening today. As will be described in detail in the following sections, much of the automatic organization and annotation for personal photos can be achieved by recording context information at the time of photo capture (such as time and location) in conjunction with content-based image analysis techniques to detect, for example, the existence of faces and buildings in the photos. These features, although promising, require further research and development to be able to deliver robust performance in real-life photo applications. At this point other semi-automatic or work-around schemes to leverage a user's manual annotation input have been proposed. For example, an initial annotation by the user at the time of photo capture can then be used later by the system to suggest more annotation options to the user (Wilhelm et al., 2004); a user does manual bulk annotation of faces appearing in a group of photos, after which the system automatically assigns the annotation to the faces in each photo (Zhang et al., 2004); and how a system could motivate its users to do enjoyable annotation is also considered (Kustanowitz & Shneiderman, 2004).

As seen in this section, photo management applications have been developed incrementally adding feature after feature whenever a new technology allowed it. In drawing up the last row in Figure 1 for a truly online automatic photo-sharing application for personal photo management, the role of mobile devices should be considered carefully especially in the light of the upsurge in the use of camera phones.

From this progression, we can now envisage a scenario in which a user takes photos at a party with her camera phone, these are instantly uploaded to an online photo server where they are fully and automatically processed and annotated, then a few days later, the user visits her uncle's house and searches on her phone for those photos taken at the party, and shares them by passing her phone around or by playing a slide show on the TV screen or an interactive wall in the house.

DESIGNING MOBILE INTERFACE FOR PHOTO MANAGEMENT

From the foregoing discussions we can see that personal photo management with a mobile device could involve activities such as:

- Capturing (taking photos)
- Storing and/or uploading
- Organizing and annotating
- Browsing and searching
- Sending and sharing of personal photos

However, not all of these activities need to be conducted on the mobile device itself. We need to consider ways in which a mobile device is best used in conjunction with other technology in varying degrees of division and overlap of task. In considering *mobile information ecologies* (Jones & Marsden, 2006, pp. 280-286) such as how a mobile device's usage should fit with other devices, physical resources, network availability and other context sources should be considered. For example, where processing power is not sufficient on the mobile device which captured a photo, uploading to a server which can index the photo more quickly and then send the result back to the mobile device could be a better solution. It is often more convenient to enter long textual descriptions for a photo using a desktop PC when the user returns home, while a short annotation at the time of capture might be still useful for facts that could have otherwise been forgotten by the time the user returns home. Viewing slide shows of photos at a gathering of relatives may be best served with a TV screen rather than viewing on the mobile device.

Depending on how a particular service or product has been designed in its use of resources in the chain of activities from capturing to storing to searching to sharing with friends, the user may need to interact with different types of interfaces, or this could be to some extent transparent and hidden from the user. For example, photos stored on a remote server or on somebody else's mobile device (in the case of peer-to-peer architecture in resource allocation) could be downloaded to the user's mobile device in the background while the user is browsing photos without them having to use a separate interface to download or browse photos stored on the server. In either case, the often quoted problems of user-interface design for mobile devices seem to remain true for personal photo management. These include details such as;

- Limited screen space
- Limited input mechanism especially awkward text input
- Potentially distracting usage environments

Many interaction and visualization related issues are raised due to these limitations. Unfortunately, we have to live with them because these natural limitations arise from the fact that most mobile devices need to be, by definition, small and mobile. Studies on Web page searching on mobile devices (Jones et al., 1999; Jones et al., 2003), though not specifically on photo searching, form a useful starting point for searching within a mobile photo management context; ideas for visualization on mobile devices have been proposed, especially for displaying interactive maps for location-based navigation (Chittaro, 2006).

The minimal attention user interface (MAUI) (Pascoe et al., 2000) tries to design a mobile interface that requires minimal user attention. This is especially intended to assist field workers who use a PDA during their physically demanding tasks. In a similar vein, use of "push" technology has been proposed to reduce the amount of user interaction on a mobile device by shifting the user's interaction burden to background processing by the device (or the server the device is connected to). For example, if the device can predict which photos the user wants to see on the mobile screen at this moment, the system can display those photos without the user needing to make the effort to enter the search query, possibly delivering the relevant photos as soon as the user turns on the device. How to accurately predict which photos the user will want to view at a given time is of course the main challenge of such an approach. Examples of such technologies include the use of collaborative filtering with the data collected from explicit preference indications (Gurrin et al., 2003), and use of attention data collected from the user's daily web browser log data (Gurrin et al., 2006). The key point of this approach is to reduce the frequency and amount of the user's interactions with the device. We will see an example of how a small number of selective personal photos are displayed before a user's query input on the first browsing screen in the next section.

The following are some of the photo browsing techniques that could be suitable for a mobile interface:

- **Thumbnail browsing:** Spatially presenting multiple miniaturized photos allows easy browsing that leverages the efficient human visual system, and has been used in almost all desktop photo management systems. Although the screen is much smaller on a mobile device, it is still a useful technique and widely used in mobile interfaces for photo browsing.
- **Smart thumbnail view** (Wang et al., 2003)**:** A photo usually contains a main focus of interest (for example, the face of a friend) as well as unnecessary visual elements (for example, strangers in the background or a large background area). By automatically determining the "regions of interest" within a photo, the interface can crop the photo to show only the area that is pertinent to the viewer. By identifying multiple regions of interest in a photo, the interface can guide the user, automatically moving the view window over a photo from one region to another in the order of importance of the regions.
- **Rapid, serial, visual presentation (RSVP)** (De Bruijn & Spence, 2000)**:** Temporally presenting multiple photos one by one as in a slide show seems particularly suitable for a small screen (De Bruijn, Spence & Chong, 2002), although having to keep focusing on the flipping-through of the images requires continuous user attention, and is thus a disadvantage when using a device with such an interface if the user needs to check their environment frequently (for example, while walking or waiting for a bus).
- **Speed Dependent automatic zooming (SDAZ):** When scrolling a page, the photos become smaller (zoomed-out) showing more of them, while it is zoomed-in when scrolling speed is reduced or ceases. This technique attempts to use the context of screen browsing. Some variations of this idea have been evaluated on a mobile device with promising results (Patel et al., 2004).
- **Key photo selection:** When multiple photos need to be displayed on a small screen, the system can determine one of those photos that is most representative of the photos and simply show the one chosen photo (while indicating that there are more to be viewed if desired). In this way, the screen space is saved and the user can browse more photos one by one if they wish to. Of course, selecting one representative photo from a group of photos taken at a particular event is an interesting research question in itself.

Some of these techniques were originally developed for desktop interfaces while others originated in PDA interfaces. These techniques and variants of them are currently being investigated for mobile interfaces. We can expect to see some of these techniques appearing in mobile photo management applications in the near future. Two strong features for enhancing photo management on mobile devices come from the technology camp, and are the subject of the remainder of this chapter. These are:

- **Content-based image analysis:** Computer vision techniques can be used to analyze the image content to classify, label, or identify something meaningful in photos for searching and browsing.
- **Context-awareness:** Context such as time, location (from GPS) and people present at the time of capture can be recorded and used to enhance metadata for searching and browsing.

As we will see in the following section, leveraging these technical elements can significantly enhance mobile interaction for photo management by enriching metadata. They can also be used subsequently to derive other useful metadata, which can in turn reduce the user's photo organization and annotation effort, and possibly enable the use of simple yet powerful time- and map-based interfaces suitable for mobile devices.

ENHANCING INTERACTION FOR MOBILE PHOTO MANAGEMENT WITH CONTENT AND CONTEXT

Content-Based Image Analysis

The use of content-based image analysis techniques for indexing and retrieving images has been an active area of research in the field of computer vision and information retrieval for many years and is neatly summarized in Smeulders et al., (2000), although even in the intervening years there have been further developments. Current approaches to content-based image retrieval can broadly be divided into three different approaches, namely using low-level features, using high-level semantic features and using segmented objects, which we now describe in turn.

Analyzing visual features of an image into low-level features such as color, shape, and texture has been the major building block for indexing image databases in order to classify and retrieve images in terms of their visual characteristics. This approach can be characterized as computationally efficient and undemanding, since these image features can be identified directly from the encoded (compressed) form of the images. Similarity between the low-level features of images can be computed simply based on intersecting histograms representing color or texture bands, where these histograms are derived for the entire image or for regions within the image. While they are computationally efficient and scalable to replicate, low-level representations of images most often do not correspond to high-level, semantic concepts that humans use when we see images. We can say that color, texture and shape are a crude first approximation to semantic image content, but very often they do not satisfy our requirement for recognizing, understanding, searching and browsing images. This difference between what low-level features offer, and what users require, is known in the literature as the "semantic gap," and has been a difficult research problem to tackle (Har et al., 2006; Smeulders et al., 2000).

The second general approach to image retrieval addresses the semantic gap head-on by trying to automatically detect semantic units directly from image content. Such semantic concepts can include almost anything, but generic concepts such as faces, buildings, indoor/outdoor, and landscape/cityscape are often used because they give general applicability. The set of possible semantic features we could detect is influenced by the use that individual detected features can offer, and as this is mostly in image classification and image retrieval, the set of possible features which we *could* calculate is enormous. In order to provide some structure and to limit the set of semantic features to use in image retrieval, we usually arrange features into an ontology which is a hierarchical arrangement of semantic topics, like the LSCOM ontology (Naphade et al., 2006). The LSCOM ontology has just under 1,000 concepts taken from the domain of broadcast TV news, but most of these concepts could be applied to any visual media, including personal photos.

The main challenge with using semantic features in applications such as personal photo management is in building classifiers to automatically detect the features. Semantic features are usually detected based on an analysis of low-level features like color, texture and shape. They are usually constructed by using a machine learning algorithm to *learn* the presence and absence of features associated with individual semantic concepts based on some training set. In the early days of using semantic features where the number of features was of the order of dozens, this was a scalable approach, but as we move towards detecting several hundred features or more, then the approach of building and training individual feature detectors does not scale, and this is one of the main challenges facing the field currently.

A second major challenge in automatically detecting features is improving the accuracy and reliability of the feature detection. Performance assessment of feature detection is carried out as part of the annual TRECVid evaluation benchmarking campaign[7] in which many (70+) participating research groups from around the world benchmark the performance of their systems for automatically detecting high-level concepts appearing in video sequences. In particular, and what makes this activity relevant to content-based image retrieval tasks such as photo managements, is that TRECVid feature detection is mostly based on shot keyframes, which are still images taken from within video shots[8]. At TRECVid 2006 benchmarking the performance of only 39 feature detectors including the presence of buildings, desert, roads, faces, animals, airplanes, cars, explosions and so on, was a significant activity for the participating groups. It is believed that building semantic feature detectors which depend upon each other, in the same way that concepts in the LSCOM ontology are arranged in a hierarchical dependency, will lead to improved feature detection accuracy (Naphade et al., 2002; Wu et al., 2004).

The final approach to content-based image retrieval that we will mention is to detect, and then use, *objects* that appear in an image as the basis for retrieval. In the approaches described so far, the processing is done on the entire image whereas in this approach we seek to identify and segment the major objects that appear within an image and to use them, rather than the whole frame, for retrieval. For example if we seek to find photos of boats then we can use a segmented image of a boat object taken from an image, independent of the background, and retrieve other objects from a photo collection based on their color, texture and/or shape. As an example of this, Sav et al. (2006) describe a system to allow manual segmentation of semantic objects from query images which are then matched against segmented objects in database images. A similar approach, albeit applied to video rather than to image retrieval, is reported by Sivic et al. (2006) used for the Google Video Search Engine[9]. What these, and a number of object-based retrieval applications which are experimental in nature, have in common is that they use segmented objects as the basis for retrieval, yet the task of automatic or semi-automatic image segmentation remains one of the most challenging image processing tasks and represents a significant hurdle towards making object-based image retrieval more widespread.

Applying content-based analysis methods of these types can enable the use of photos as queries to search for similar ones or objects within one photo to find similar objects in other photos. However, as is made clear from the above, these technologies are either rather unreliable, since they lack the power to capture perceived semantic features because they are too low-level, such as the color-based features, or they are rather specialized, in the form of learned features of particular objects. These methods thus have considerable possible utility, but are not, at present at any rate, suitable for robust and reliable photo management, but they do offer interesting potential when used in combination with context features associated with photo capture as explored in the next section.

Context-Awareness

Context information recorded at the time of photo capture can be used to assist with photo management on a mobile device in a number of ways. Particularly important given the high volume of photos often taken with devices such as camera phones and the interaction issues reviewed earlier in this chapter, a key feature in the use of the context of photo capture is that it can generally be used entirely automatically. In this section we examine the following easily captured and used context features: time, location, lighting levels and weather.

Time

Chronology is one of the most important clues when a user is looking for photos (Rodden, 2002). We know that users often remember at least the rough date/time of photo capture, even if they cannot remember exact details. Digital cameras routinely record the time of photo capture in the EXIF header of photos captured. This data allows photos to be indexed using fields such as year, month, day of the month, day of the week, hour of the day. In addition it is possible to derive more descriptive fields, such as season, weekday, or weekend, which will further aid user interactions. Indexing by time along multiple dimensions like this is useful when the user only remembers certain facets of the temporal context surrounding photo capture. For example, they may remember only that a photo was taken in the summer, in the evening, on a certain day of the week, or on the weekend.

Location

The integration of a location capture device and a camera provides the ideal scenario for location stamping of digital photo collections. However, at this point consumer digital cameras do not have integrated location stamping capabilities. While awaiting the arrival to the market of cameras which incorporate this capability, it is possible to utilize a separate GPS device to record the location at which photos are taken, and then via a timestamp matching process, incorporate the location of photo capture into the EXIF header. In our own work, we capture the locations using a small portable GPS device tracklog stored every 10 seconds, and utilize this tracklog in the location stamping process. In order to map raw GPS coordinates to real world locations, we utilize a gazetteer which typically allows the indexing of each photo at three separate levels: country, city and state, and town. The level and accuracy of location stamping depends on the granularity of the available gazetteers.

The key benefits of labeling digital photos with their location are that it enables us to support a number of access methodologies: search by actual location (country, city,town, even street), search by proximity to a location, or by proximity to other photos. By using such information the browsing space (number of photos that a user has to browse through) when seeking a particular photo can be drastically reduced.

In addition, it is possible to present a user with a map-based interface to their photo collection, with photos, or icons, plotted on a map. For example the Microsoft WWMX system (Toyama et al., 2003) takes this approach, while Google Maps[10] allows its map-specific APIs to be easily incorporated into a Web-based personal photo application thus saving development effort.

Previous research (Gurrin et al., 2005) shows that the integration of location context into a time-context based system reduces mean time to locate a given photo within an experimental collection of 8,000 photos from 32 seconds to 18 seconds, and reduces the mean number of query iterations required to locate the given photo from 3.7 to 2.8.

Other Context Issues

However powerful time and location are individually at supporting user search of digital photo collections, by combining these two contextual features, one can derive additional contextual features, such as *lighting levels* and *weather*. Standard astronomical algorithms (Meeus, 1999) allow us to calculate the environmental lighting level at the time and location of photo capture. A photo taken at 10 a.m. will be in daylight in most parts of the world, but this is not always the case, for example, in parts of Scandinavia and similar high-latitude locations this time could signify dawn, or even darkness, depending on the time of year. We use astronomical algorithms to calculate sunrise and sunset times for any location on any date, and using these algorithms we can associate a daylight status (daylight, darkness, dawn or dusk) with each photo based on its time and GPS location of capture, and thereby automatically annotate each photo with this information. When searching for a photo it is probably more likely that a user will remember that it was dark when a particular photo was being taken, than the exact time that they took the picture.

Another feature that can be used to annotate each photo is the prevailing weather conditions. There are 10,500 international weather stations dotted all across the globe which log weather data a number of times each day. Given this information, and readily available access to the weather data logs via the Web, one can annotate each photo with the weather data (clear, cloudy, rainy, or snowy) from the closest international weather station at the time the photo was taken.

Finally, people present at the time of photo capture could be yet another potentially useful context that can be captured. By using a Bluetooth device, people nearby who have Bluetooth-enabled devices can be picked up and recorded, and this information can complement other methods such as face recognition (Davis et al., 2005) effectively combining context with content-based techniques.

Content-based analysis and context-awareness as discussed so far can be applied to user access to photo collections via a mobile device to significantly enhance the user interaction on such a device. The next section introduces a prototype of such a system under development in our laboratory.

Mobile Photo Access: An Example

The MediAssist mobile interface (Gurrin et al., 2005) to personal digital photo libraries has been designed to minimize user input and proactively recommend photos to the user. Consequently, it supports the following three access methodologies from a mobile device:

- ‘**My Favorites’:** The first screen a user sees when accessing their archive using a mobile device (see Figure 2a) is a personalized thumbnail listing of the top 10 most popular photos based on a user’s history of viewing full-size photos, where this history data is gathered both from mobile devices and conventional desktop device access.
- **Search functionality:** Primarily based on location and the derived annotations. The aim is to reduce the level of user interaction required to quickly locate relevant content. In order to maximize screen real-estate available for browsing the photo archives, search options are hidden in a panel that slides into view when a user wants to search (see Figure 2b) and then disappears afterwards until required again (Figure 2a).
- **Browsing the collection by events:** Even by supporting the two access methods on a mobile device a user may still end up having to spend time scrolling through screens of photos if many were taken at the same time and place. To address this issue, the interface presents results to the user clustered into *events* and ordered by date and time. Events are logical combinations of photos taken in close proximity of location and time. Event clustering of photos can either be a rule-based process (e.g., no photos taken for a period of 90 minutes signifies the end of an event), or a clustering process where photos are grouped together based on location and/or time and the unique clusters extracted to comprise events in a personal photo collection.

A user accessing the photo archive is immediately presented with the ‘My Favorites’ screen, of their most accessed photos, helping to reduce user interaction. If the required photo is not in the favorites, the user engages in a process of searching, followed by browsing of the search results, so in effect it is a two-phase search. The search options are: three level location (country, state, and city or town), season, weather, and lighting status, as shown on the sliding panel in Figure 2b. The contents of the location drop-down boxes are personalized to the user’s collection to minimize user input. Season, weather, and lighting status are included to filter the search results thereby reducing the amount of browsing effort required to locate the desired photos.

Figure 2. MediAssist mobile interface takes advantage of context information to automatically organize personal photos

(a) 'My favorite photos' with search panel down

(b) Searching the archive with a representative photo results as event summary

Simply presenting a (potentially long) list of photo thumbnails in response to a query is not an ideal interaction scenario for the user of a mobile device. A more 'mobile friendly' technique is to group photos into events and using a single 'key' thumbnail which represents an event, as shown in Figure 2b (one event displayed). This is done by automatically grouping photos together into logical sets by examining when and where clusters of photos co-occur and choosing a single representative photo to represent the whole cluster. Typically the photo chosen to represent the cluster is the middle photo from a temporally organized listing of the cluster photos. In future work we will focus on judiciously choosing the most representative photo in a query-biased manner taking account of context and content data associated with photos in the event. These clusters are then presented to the user, ordered by time and date. Tapping on a thumbnail photo on screen presents the user with a full-screen photo, and it is this detailed viewing of a photo that is used to support the 'My Favorites' access method. Associated with each thumbnail is a small arrow button on the right side of the thumbnail. Tapping on this arrow brings the user to a screen showing all photos from that particular event, once again organized by date and time. In an experiment, the broad context searching capabilities of the mobile MediAssist system were shown to clearly outperform a more conventional time-only based system (Gurrin et al., 2005).

CONCLUSION

Interesting avenues for application scenarios are already appearing in literature which leverage context and/or content analysis for mobile photo annotation and searching. A mobile photo management system (Sarvas et al., 2004) records location, time and user data at the time of photo capture and then compares this with other already annotated metadata from other users and presents an inferred annotation for the new photo to the user. The Photo-to-Search system (Fan et al., 2005) allows a user to take a photo with a camera phone, the system then searches for visually similar images from the web and returns the result on the mobile device. While these make interesting applications and their evaluation with users will be highly important for developing future mobile systems, we need further application ideas and testing in other tasks of photo management to be able to explore more diverse application possibilities and new kinds of functions.

As we have seen in this chapter, as the number of digital photos each person needs to manage in their collection continues to grow, it will be inevitable that some form of automatic management is used as an integral module of photo management systems to help the user cope with the number of photos, even if running only at the background. Leveraging context data at the time of photo capture and use of steadily improving content-based image analysis will form a crucial part in making automatic organization of personal photos feasible. The role of mobile devices such as camera phones in this application area is also growing very rapidly as such devices become more powerful technically and more ubiquitous and more accepted socially. However, a mobile device will not, in itself, be designed to do every task of photo management. They will be designed to be optimally used in conjunction with other devices such as desktop PCs and laptops, TVs, and other information appliances, depending on technical and social situations. Automatic organization, then, is a vital back-end technology in the chain of personal photo management tasks of which mobile interaction is a part.

ACKNOWLEDGMENT

We gratefully acknowledge support from the aceMedia Project under contract FP6-001765, Science Foundation Ireland under grant number 03/IN.3/I361, and Enterprise Ireland under grant number CFTD-03-216.

REFERENCES

Chittaro, L. (2006). Visualizing information on mobile devices. *IEEE Computer*, *39*(3), 40–45.

Davis, M., Smith, M., Canny, J., Good, N., King, S., & Janakiraman, R. (2005). Towards context-aware face recognition. In *Proceedings of the 13th Annual ACM International Conference on Multimedia '05* (pp. 483-486). Singapore

De Bruijn, O., & Spence, R. (2000). Rapid Serial Visual Presentation: a Space-Time Trade-off in Information Presentation. In *Proceedings of the Working Conference on Advanced Visual Interfaces AVI '00* (pp.189-192). Palermo, Italy.

De Bruijn, O. Spence, R., & Chong, M. Y. (2002). RSVP browser: Web browsing on small screen devices. *Personal and Ubiquitous Computing, 6*(4), 245-252.

Enser, P. (1995). Pictorial information retrieval. *Journal of Documentation*, *51*(2), 126-170.

Fan, X., Xie, X., Li, Z., Li, M., & Ma, W.-Y. (2005). Photo-to-search: Using multimodal queries to search the web from mobile devices. In *MIR '05 - Proceedings of the 7th ACM SIGMM International Workshop on Multimedia Information Retrieval* (pp. 143–150). Singapore.

Frohlich, D., Kuchinsky, A., Pering, C., Don, A., & Ariss, S. (2002). Requirements for photoware. In *CSCW'02 -Proceedings of the 2002 ACM Conference on Computer Supported Cooperative Work* (pp. 166–175). New Orleans, LA.

Gurrin, C., Brenna, L., Zagorodnov, D., Lee, H., Smeaton, A. F., & Jahansen, D. (2006). Supporting mobile access to digital video archives without requiring user queries. In *MobileHCI '06 - Proceedings of the 8th International Conference on Human-Computer Interaction with Mobile Devices and Services* (pp. 165-168). Espoo, Finland.

Gurrin, C., Jones, G., Lee, H., O'Hare, N., Smeaton, A. F., & Murphy., N. (2005). Mobile Access to Personal Digital Photograph Archives. In *MobileHCI'05 - Proceedings of the 7th International Conference on Human Computer Interaction with Mobile Devices and Services* (pp. 311-314). Salzburg, Austria.

Gurrin, C., Smeaton, A. F., Lee, H., Donald, K. M., Murphy, N., O'Connor, N., & Marlow, S. (2003). Mobile access to the Físchlár-news archive. In F. Crestani, M. Dunlop, & S. Mizzaro (Eds.), In *Mo-*

bileHCI'03 – Proceedings of the 5th International Symposium on Human Computer Interaction with Mobile Devices and Services, Workshop on Mobile and Ubiquitous Information Access (LNCS Vol. 2954) (pp. 124-142), Berlin/Heidelberg, Germany: Springer.

Har, J. S., Lewis, P. H., Enser, P., & Sandom, C. J. (2006, January15-19). Mind the Gap: Another Look at the Problem of the Semantic Gap in Image Retrieval. In *Proceedings of the SPIE Multimedia Content Analysis, Management, and Retrieval, 6073(1)*. San Jose, CA.

Jones, M., Buchanan, G., & Thimbleby, H. (2003). Improving web search on small screen devices. *Interacting with Computers*, *15*(4), 479-495.

Jones, M., & Marsden, G. (2006). *Mobile interaction design*. Wiley.

Jones, M., Marsden, G., Mohd-Nasir, N., & Boone, K. (1999). Improving Web Interaction on Small Displays. In *Proceedings of the 8th World Wide Web Conference* (pp. 1129-1137). Toronto, Canada.

Kindberg, T., Spasojevic, M., Fleck, R., & Sellen, A. (2005). The ubiquitous camera: An in-depth study of camera phone use. *IEEE Pervasive Computing*, *4*(2), 42-50.

Kustanowitz, J., & Shneiderman, B. (2004). Motivating Annotation for Digital Photographs: Lowering Barriers While Raising Incentives. In *Technical Report, HCIL-2004-18, CS-TR-4656, ISR-TR-2005-55, HCIL University of Maryland, 2004*.

Meeus, J. (1999). *Astronomical Algorithms, 2nd ed*. Willmann-Bell.

Naphade, M., Kozintsev, I., & Huang, T. (2002). A factor graph framework for semantic video indexing. *IEEE Transactions on Circuits and Systems for Video Technology*, *12*(1), 40-52.

Naphade, M., Smith, J. R., Tesic, J., Chang, S.-F., Hsu, W., Kennedy, L. et al.(2006). Large-scale concept ontology for multimedia. *IEEE Multimedia*, *13*(2), 86-91.

O'Hare, N., Lee, H., Cooray, S., Gurrin, C., Jones, G., & Malobabic, J. (2006). MediAssist: Using content-based analysis and context to manage personal photo collections. In H. Sundaram, M. R. Naphade, J. R. Smith, & Y. Rui: (Eds.), *CIVR2006 - 5th International Conference on Image and Video Retrieval* (Vol. LNCS 4071) (pp. 529-532). Tempe, AZ.

Palen, L., & Salzman, M. (2002). Beyond the handset: Designing for wireless communications usability. *ACM Transactions on Computer-Human Interaction*, *9*(2), 125–151.

Pascoe, J., Ryan, N., & Morse, D. (2000). Using while moving: HCI issues in fieldwork environments. *ACM Transactions on Computer-Human Interaction*, *7*(3), 417–437.

Patel, D., Marsden, G., Jones, S., & Jones, M. (2004). An Evaluation of Techniques for Browsing Photograph Collections on Small Displays. Lecture Notes in Computer Science (LNCS). In *MobileHCI'04 - Proceedings of the 6th International Conference on Human-Computer Interaction with Mobile Devices and Services* (pp. 132-143). Glasgow, Scotland.

Perry, M., O'hara, K., Sellen, A., Brown, B., & Harper, R. (2001). Dealing with mobility: Understanding access anytime, anywhere. *ACM Transactions on Computer-Human Interaction*, *8*(4), 323–347.

Rodden, K. (1999). How do people organise their photographs? In *IRSG'99 - Proceedings of the BCS IRSG 21st Annual Colloquium on Information Retrieval Research*, Glasgow, Scotland.

Rodden, K. (2002). Evaluating Similarity-based Visualisations as Interfaces for Image Browsing. In *University of Cambridge Technical Report, UCAM-CL-TR-543*.

Rodden, K., & Wood, K. (2003). How do people manage their digital photographs? In *CHI'03 - Proceedings of CHI '03: Proceedings of the SIGCHI conference on Human factors in computing systems*, (pp. 409-416). Fort Lauderdale.

Sarvas, R., Herrarte, E., Wilhelm, A., & Davis, M. (2004). Metadata Creation System for Mobile Images. In *MobiSys '04 - Proceedings of the 2nd International Conference on Mobile Systems, Applications, and Services* (pp. 36–48). Boston, MA.

Sarvas, R., Oulasvirta, A. and Jacucci, G. (2005). Building social discourse around mobile photos: a systematic perspective. In *MobileHCI' 05 – Proceedings of the 7th International Conference on Human Computer Interaction with Mobile Devices and Services* (pp. 31-38). Salzburg, Austria

Sav, S., J. F. Jones, G., Lee, H., O'Connor, N. E., & Smeaton, A. F. (2006). Interactive experiments in object-based retrieval. In H. Sundaram, M. R. Naphade, J. R. Smith, & Y. Rui: (Eds.), *CIVR 2006 – Proceedings of the 5th International Conference on Image and Video Retrieval* (LNCS 4071). Tempe, AZ.

Sivic, J., Schaffalitzky, F., & Zisserman, A. (2006). Object level grouping for video shots. *International Journal of Computer Vision*, *67*(2), 189–210.

Smeulders, A. W. M., Worring, M., Santini, S., Gupta, A., & Jain, R. (2000). Content-based image retrieval at the end of the early years. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *22*(12), 1349–1380.

Toyama, K., Logan, R., Roseway, A., & Anandan, P. (2003). Geographic location tags on digital images. In *MULTIMEDIA'03 - Proceedings of the 11th ACM International Conference on Multimedia* (pp. 156-166). Berkeley, CA.

Van House, N., Davis, M., Ames, M., Finn, M., & Viswanathan, V. (2005). The uses of personal networked digital imaging: an empirical study of cameraphone photos and sharing. In *CHI'05 - Proceedings of the CHI 2005 Extended Abstracts on Human Factors in Computing Systems* (pp. 1853-1856). Portland, OR.

Wang, M.-Y., Xie, X., Ma, W.-Y., & Zhang, H.-J. (2003). MobiPicture: Browsing Pictures on Mobile Devices. In *MULTIMEDIA'03 - Proceedings of the 11th ACM International Conference on Multimedia* (pp.106–107). Berkeley, CA.

Wilhelm, A., Takhteyev, Y., Sarvas, R., House, N. V., & Davis, M. (2004). Photo annotation on a camera phone. In *CHI'04 Extended Abstracts on Human Factors in Computing Systems* (pp.1403-1406). Vienna, Austria.

Wu, Y., Tseng, B., & Smith, J.R. (2004). Ontology-based multi-classification learning for video concept detection. In *ICME 2004 - Proceedings of the IEEE International Conference on Multimedia and Expo* (pp. 1003-1006). Taipei, Taiwan.

Zhang, L., Hu, Y., Li, M., Ma, W., & Zhang, H. (2004). Efficient propagation for face annotation in family albums. In *MULTIMEDIA'04 - Proceedings of the 12th annual ACM international conference on Multimedia* (pp. 716-723). New York, NY.

KEY TERMS

Blog: A type of website in which the user adds regular written contributions on his/her own life or thoughts, as in a journal or diary. Contracted from *weblog,* usually the entries are in reverse chronological order, and readers are allowed to add their own comments.

Context-Awareness: A system that can use information about the circumstances under which it is being used. For example a context-aware device will use the current time and location where it is being used to infer what would be the most beneficial piece of information to display for the user.

Content-Based Image Retrieval (CBIR): An application of computer vision to image retrieval, in which an image's content (its color, texture, shapes, objects or faces in it, etc.) is automatically analyzed to index the image for subsequent retrieval.

Information Retrieval (IR): An interdisciplinary field of study that deals with searching for information in documents (papers, books, pictures, video clips, or any other item that contain useful information). IR systems seek to return to users documents which satisfy their current information need as expressed through some form of search request which may comprise components in one or more media.

Global Positioning System (GPS): A satellite navigation system in which more than two dozen satellites broadcast precise timing signals by radio, allowing any GPS receiver device to accurately determine its location.

Photoware: A software application used for personal photo management. Although the term emerged when online sharing of photos became common in personal photo management software, in this chapter we use the term in a more general sense.

Web 2.0: The second generation of Internet-based services in which the Web itself is a platform for users to directly use and share information on the Web, often characterized by its highly-dynamic and highly-interactive Web interfaces and pulling together the distributed resources from independent developers of contents.

ENDNOTES

1 Flickr. Available at http://www.flickr.com/ (Retrieved January 2007)

2 Adobe Photoshop Album. Available at http://www.adobe.com/products/photoshopalbum/starter.html (Retrieved January 2007)

3 ACDSee. Available at http://www.acdsee-guide.com/ (Retrieved January 2007)

4 Google Picasa. Available at http://picasa.google.com/ (Retrieved January 2007)

5 Flickr. Available at http://www.flickr.com/ (Retrieved January 2007)

6 Yahoo! 360. Available at http://360.yahoo.com/ (Retrieved January 2007)

7 TRECVid 2006 Guideline. Available at: http://www-nlpir.nist.gov/projects/tv2006/tv2006.html (Retrieved January 2007)

8 A "video shot" is an unbroken sequence of frames taken by a single camera. Shot boundaries occur at camera changes.

9 The Google Video Search Engine. Available at: http://video.google.com/ (Retrieved January 2007)

10 Google Maps. Available at http://maps.google.com/ (Retrieved January 2007)

Chapter VI
Understanding One-Handed Use of Mobile Devices

Amy K. Karlson
University of Maryland, USA

Benjamin B. Bederson
University of Maryland, USA

Jose L. Contreras-Vidal
University of Maryland, USA

ABSTRACT

Mobile phones are poised to be the world's most pervasive technology, already outnumbering land lines, personal computers, and even people in some counties. Unfortunately, solutions to address the usability challenges of using devices on the move have not progressed as quickly as the technology or user distribution. Our work specifically considers situations in which a mobile user may have only one hand available to operate a device. To both motivate and offer recommendations for one-handed mobile design, we have conducted three foundational studies: a field study to capture how users currently operate devices; a survey to record user preference for the number of hands used for a variety of mobile tasks, and an empirical evaluation to understand how device size, interaction location, and movement direction influence thumb agility. In this chapter we describe these studies, their results, and implications for mobile device design.

INTRODUCTION

The handheld market is growing at a tremendous rate; the technology is advancing rapidly and experts project that over one billion mobile phones will be sold in this year (2007) alone (Milanesi et al., 2007). To meet customer demand for portability and style, device manufacturers continually introduce smaller, sleeker profiles to the market. Yet advances in battery power, processing speed, and memory allow these devices to come equipped with increasing numbers of functions, features, and applications. Unfortunately these divergent trends are at direct odds with usability: richer content accessed through shrinking input and output channels simply makes devices harder to use. The unique requirements for mobile computing only compound the problem, since mobile use scenarios can involve unstable environments, eyes-free interaction, competition for users' attention, and varying hand availability (Pascoe, Ryan, & Mores, 2000). While each of these constraints requires attention in design, we are currently interested in issues of usability when a user only has only one hand available to operate a mobile device.

Devices that accommodate single-handed interaction can offer a significant benefit to users by freeing a hand for the host of physical and mental demands common to mobile activities. But there is little evidence that current devices are designed with this goal in mind. Small, light mobile phones that are easy to control with one hand are unfriendly to thumbs due to small buttons and crowded keypads. Larger devices, such as personal digital assistants (PDAs) are not only harder to manage with a single hand, they tend to feature more (rather than larger) buttons, as well as stylus-based touchscreens whose rich interface designs emphasize rich information content, but often offer targets too small, and/or too distant, for effective thumb interaction.

While it may seem obvious which features inhibit single-handed use, there has been relatively little systematic study of enabling technologies and interaction techniques. Most commercial and research efforts in one-handed device interaction have focused primarily on either a specific technology or task. For example, accelerometers have been explored to support tilt as a general input channel for handheld devices (Dong, Watters, & Duffy, 2005; Hinckley, Pierce, Sinclair et al., 2000; Rekimoto, 1996), while media control (Apple, 2006; Pirhonen, Brewster, & Holguin, 2002) and text entry (Wigdor & Balakrishnan, 2003) have been popular tasks to consider for one-handed device operation. But in the varied landscape of mobile devices and applications, one-handed design solutions must ultimately extend to a wide range of forms and functions. We began our investigation of this problem by looking at the fundamental human factors involved in operating a device with a single hand.

In this chapter, we report on three studies conducted to understand different aspects of one-handed mobile design requirements. We first ran a field study to capture the extent to which single-handed use is currently showing up "in the wild." Second, we polled users directly to record personal accounts of current and preferred device usage patterns. The results from these studies help motivate one-handed interface research, and offer insight into the devices and tasks for which one-handed techniques would be most welcomed. Finally, we performed an empirical evaluation of thumb tap speed to understand how device size, target location, and movement direction influence performance. From these results we suggest hardware-independent design guidelines for the placement of interaction objects. Together our findings offer foundational knowledge in user behavior, preference, and motor movement for future research in single-handed mobile design.

BACKGROUND

The physical and attention demands of mobile device use were reported early on for fieldworkers (Kristoffersen & Ljungberg, 1999; Pascoe et al., 2000), from which design recommendations for minimal-attention and one-handed touchscreen interface designs emerged (Pascoe et al., 2000). Though well suited to the directed tasks of fieldwork, the guidelines do not generalize to the varied and complex personal information management tasks of today's average user. Research of the effects that mobility has on attention and user performance continues (Oulasvirta, Tamminen, Roto et al., 2005), as well as how these factors can be replicated for laboratory study (Barnard, Yi, Jacko, & Sears, 2005).

Several approaches for one-handed device interaction have been proposed. Limited gestures sets have been explored for mobile application control with both the thumb (Apple, 2006; Karlson, Bederson, & SanGiovanni, 2005; Pascoe et al., 2000) and index finger (Pirhonen et al., 2002), but none have specifically considered ergonomic factors. Since text entry remains the input bottleneck for mobile devices, many are working on improvements, and some targeting one-handed use. Peripheral keyboards for one-handed text entry are available, such as the Twiddler (Lyons et al., 2004), but the mobile device itself must be supported by another hand, desk or lap, which violates our definition of one-handed device control. Text entry on phone keypads is generally performed with a single thumb, but methods to improve input efficiency have focused on reducing the number of key presses required, such as T9 word prediction, rather than by improving ergonomics by optimizing button sizes, locations, or movement trajectories. Accelerometer-augmented devices allow for the device's spatial orientation to serve as an input channel, and have been shown to support one-handed panning (Dong, Watters, & Duffy, 2005), scrolling (Rekimoto, 1996), and text entry (Wigdor & Balakrishnan, 2003). However, the coarse level of control tilt offers, and the potential for confusion with the normal movements of mobile computing necessarily limit the viability of tilt for generalized input.

Scientists in the medical community have studied the biomechanics of the thumb extensively for the purposes of both reconstruction and rehabilitation. The structure of the thumb is well understood (Barmakian, 1992), but only now are scientists beginning to reliably quantify the functional capabilities of the thumb. Strength has been the traditional parameter used to assess biomechanical capabilities, and recent research has established the effect movement direction has on thumb strength (Li & Harkness, 2004). Unfortunately, only standard anatomical planes have been considered, which excludes movements toward the palm that are typical of mobile device interaction. As a complement to force capabilities, others have looked at range as a characteristic of thumb movement. Kuo, Cooley, Kaufman et al. (2004) have developed a model for the maximal 3D workspace of the thumb and Hirotaka (2003) has quantified an average angle for thumb rotation. The experimental conditions for these studies, however, do not account for constraints imposed by holding objects of varying size, such as alternative models of handheld device.

FIELD STUDY

One motivation for our research in single-handed mobile designs was our assumption that people already use devices in this manner. Since current interaction patterns, whether by preference or necessity, are predictive of future behavior, they are likely to be transferred to new devices. This suggests that designs should become more accommodating to single-handed use, rather than less, as the tradition has been. To capture current behavior, we conducted an in situ study of user interaction with mobile devices. The study targeted an airport environment for the high potential of finding mobile device users and ease of access for unobtrusive, anonymous observation.

Field Study Method

We observed 50 travelers (27 male) at Baltimore Washington International Airport's main ticketing terminal during a six hour period during peak holiday travel. Because observation was limited to areas accessible to non-ticketed passengers, seating options were scarce. We expected to observe the use of both PDAs and cell phones since travelers are likely to be coordinating transportation, catching up on work, and using mobile devices for entertainment purposes. Since most users talk on the phone with one hand, we recorded only the cell phone interactions that included keypad interaction as well. All observations were performed anonymously without any interaction with the observed.

Note that while any subject observation without consent presents a legitimate question for ethical debate, in our research we follow the federal policy on the protection of human research subjects (Department of Health and Human Services, 2005) as a guideline. The policy states that the observation of public behavior is not regulated if the anonymity of the subjects is maintained and that disclosure of the observations would not put the subjects at risk in terms of civil liability, financial standing, employability, or reputation. Since we were interested in capturing natural behavior, did not record identifying characteristics, and consider phone use while standing, walking and sitting relatively safe activities, we did not obtain subject consent.

Field Study Measures

For each user observed, we recorded sex, approximate age, and device type used: candy bar phone, flip phone, Blackberry, or PDA. A "candy bar" phone is the industry term for a traditional-style cellular phone with a rigid rectangular form, typically about 3 times longer than wide. For phone use, we recorded the hand(s) used to dial (left, right or both) and the hand(s) used to speak (left, right or both). We also noted whether users were carrying additional items, and their current activity (selected from the mutually exclusive categories: walking, standing, or sitting).

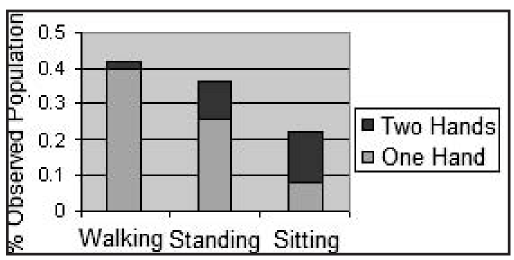
Field Study Results

Only two users were observed operating devices other than mobile phones—one used a PDA and the other a Blackberry. Both were seated and using two hands. The remainder of the discussion focuses on the 48 phone users (62.5 percent flip, 37.5 percent candy bar). Overall, 74 percent used one hand for keypad interaction. By activity, 65 percent of one handed users had a hand occupied, 54 percent were walking, 35 percent were standing, and 11 percent were sitting. Figure 1 presents the distribution of subjects who used one vs. two hands for keypad interaction, categorized by the activity they were engaged in (walking, standing, or sitting). The distribution of users engaged in the three activities reflects the airport scenario where many more people were walking or standing than sitting. It is plain from Figure 1 that the relative proportion of one handed to two handed phone users varied by activity; the vast majority of walkers used one hand, about two-thirds of standers used one hand, but seated participants tended to use two hands. However, we also recorded whether one hand was occupied during the activity, and found walkers were more likely to have one hand occupied (60 percent), followed by standers (50 percent), and finally sitters (25 percent), which may be the true reason walkers were more likely than standers to use one hand, as well as why standers were more likely than sitters to use one hand. Regardless of activity, when both hands were available for use, the percentage of one vs. two handed phone users was equal.

Analysis of Field Study

Although Figure 1 suggests a relationship between user activity and keypad interaction behavior, it is unclear whether activity influences the number of hands used, or vice versa. Furthermore, since the percentage of users with one hand occupied correlates with the distribution of one-handed use across activities, hand availability, rather than preference, may be the more influential factor

Figure 1. Airport field study: Number of hands used for keypad interaction by activity.



in the number of hands used to interact with the keypad. While use scenario certainly impacts usage patterns, the fact that users were as likely to use one hand as two hands when both hands were available suggests that preference, habit and personal comfort also play a role. Regardless of scenario, we can safely conclude that one-handed phone use is quite common, and thus is an essential consideration in mobile phone design.

Generalizability. The choice of observation location may have biased our results from those found in the general population since travelers may be more likely to be: (1) carrying additional items; (2) standing or walking; and (3) using a phone vs. PDA. Different environments, information domains, populations, and scenarios will yield unique usage patterns. Our goal was not to catalogue each possible combination, but to learn what we could from a typical in-transit scenario.

WEB SURVEY

While informative for a preliminary exploration, shortcomings of the field study were (a) a lack of knowledge about motivation for usage style; (b) the limited types of devices observed (phones); and (c) the limited tasks types observed (assumed dialing). To broaden our understanding of device use over these dimensions, we designed a Web survey to capture user perceptions of, preferences for and motivations surrounding their own device usage patterns.

Survey Method

The survey consisted of 18 questions presented on a single Web page which was accessed via an encrypted connection (SSL) from a computer science department server. An introductory message informed potential participants of the goals of the survey and assured anonymity. Notification that results would be posted for public access after the survey period was over provided the only incentive for participation. Participants were solicited from a voluntary subscription mailing list about the activities of our laboratory. In addition the solicitation was propagated to one recipient's personal mailing list, a medical informatics mailing list, and a link to the survey was posted on two undergraduate CS course Web pages.

Survey Measures

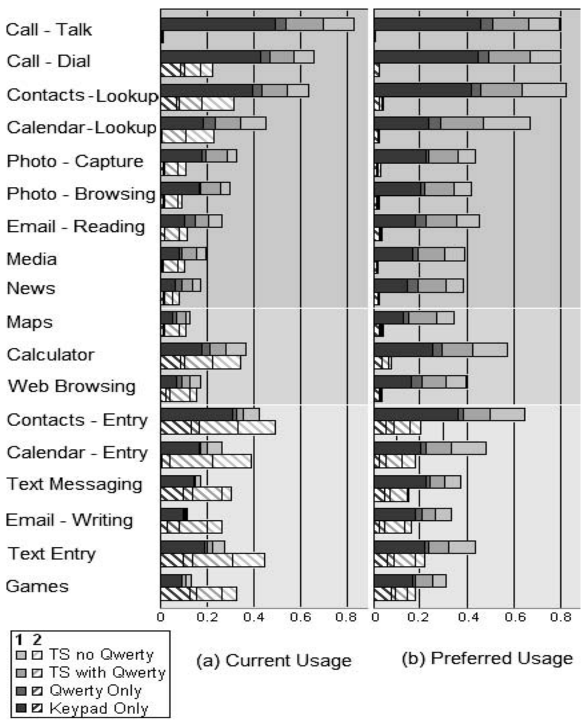
For each participant, we collected age, sex and occupation demographics. Users recorded all styles of phones and/or PDAs owned, but were asked to complete the survey with only one device in mind—the one used for the majority of information management tasks. We collected general information about the primary device, including usage frequency, input hardware, and method of text entry. We then asked a variety of questions to understand when and why people use one vs. two hands to operate a device. We asked users to record the number of hands used (one and/or two) for 18 typical mobile tasks, and then to specify the number of hands (one *or* two) they would *prefer* to use for each task. Three pairs of activities were designed to distinguish between usage patterns for different tasks within the same application, which we differentiated as "read" (e-mail reading, calendar lookup, and contact lookup) vs. "write" (e-mail writing, calendar entry, and contact entry) tasks. Users then recorded the number of hands used for the majority of device interaction and under what circumstances they chose one option over the other. Finally, users were asked how many hands they would prefer to use for the majority of interactions (including no preference), and were also asked to record additional comments.

Survey Results

Two hundred twenty-nine participants (135 male) responded to the survey solicitation. One male participant was eliminated from the remaining analysis because his handheld device was specialized for audio play only, leaving 228. The median participant age was 38.5 years. Participant occupations reflected the channels for solicitation, with 25 percent in CS, IT or engineering, 23 percent students of unstated discipline, 20 percent in the medical field, 10 percent in education, and the remainder (22 percent) from other professional disciplines.

Devices owned. The three most common devices owned were flip phones (52 percent), small candy bar phones (23 percent) and Palm devices without a Qwerty keyboard (20 percent). Palm devices with an integrated Qwerty keyboard were as common as Pocket PCs without a keyboard (14 percent). Since interaction behavior may depend on device input capabilities, we reclassified each user's primary device into one of four general categories based on the device's input channels: (1) *keypad-only (51 percent)* are devices with a 12-key numeric keypad but no touchscreen, (2) *TS-no-qwerty (23 percent)* are devices with a touchscreen but no Qwerty keyboard, (3) *TS-with-qwerty (21 percent)* are devices with a touchscreen as well as an integrated Qwerty keyboard, and finally (4) q*werty-only (5 percent)* are devices with an integrated Qwerty keyboard but no touchscreen.

Figure 2. Web survey: Number of hands (a) currently used and (b) preferred (1 hand is shown as solid, 2 hands is shown as striped) for 18 mobile tasks as a percentage of the observed population. Hand usage for each task is broken down by device type (TS = touchscreen)

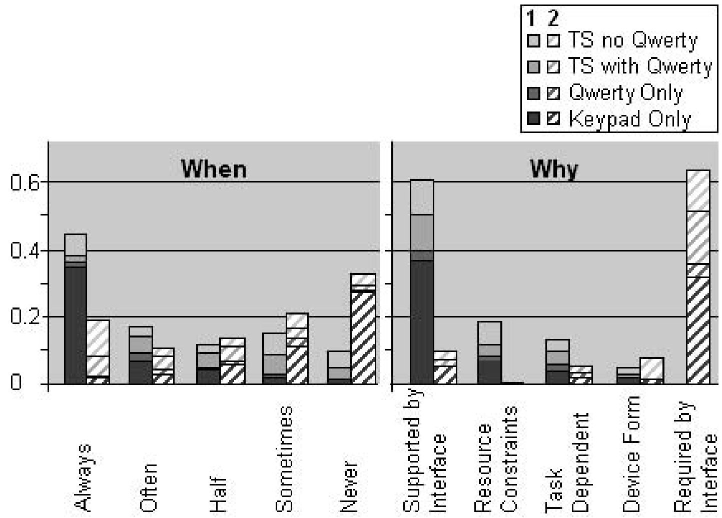


For users with multiple devices, we derived their primary device type from the text entry method reported.

Current usage patterns. Of the 18 activities users typically perform with devices, 9 were performed more often with one hand, 6 more often with two hands, and 3 were performed nearly as often with one vs. two hands. Figure 2a displays these results, with the shaded backgrounds grouping the activities by those used with one, either or two hands. Upon inspection, all of the "reading" activities were performed more often with one hand (top) and all "writing" activities with two hands (bottom). Considering users' device types, we notice that with the exception of gaming, owners of *keypad-only* devices were more likely to use one hand regardless of activity, owners of *TS-no-qwerty* were more likely to use two-hands for most activities, and those owning Qwerty based devices were more likely to use two hands when performing writing tasks, but not reading tasks.

Overall, 45 percent of participants stated they use one hand for nearly all device interactions, as opposed to only 19 percent who responded similarly for two hands. Considering device ownership, however, users of touchscreen-based devices were more likely to use two hands "always" than they were one hand (Figure 3). When participants use one hand, the majority (61 percent) perceive they do so whenever the interface supports it, the reason cited by only 10 percent of those who use two hands. Device form dictated usage behavior when the device was too small for two hands, too large for one hand, or when large devices could be supported by a surface and used with one hand. Participants cited task type as a reason for hand choice, primarily as a trade off between efficiency and resources usage: 14 percent of users selected one hand only for simple tasks (conserving resources), while 5 percent selected two hands for entering text, gaming, or otherwise for improving the speed of interaction (favoring efficiency). Finally, according to respondents, the majority of two-handed use occurs when it is the only way to accomplish the task given the interface (63 percent).

Figure 3. Web survey: The (a) frequency and (b) reasons for one (solid) and two (striped) handed device use, broken down by device type



Hand preferences. When asked how many hands users preferred to use while performing the same 18 tasks, one hand was preferred overwhelmingly to two hands for all tasks (Figure 2b). The activities with the closest margin between the number of participants who preferred one vs. two hands were playing games (13 percent) and composing e-mail (16 percent). With one exception (gaming), the activities for which more than 14 percent of users stated a preference for two hands were "writing" tasks (e.g., those that required text entry): text entry, contact entry, calendar entry, e-mail writing, and text messaging, in decreasing order. Even so, except for users of *TS-with-qwerty* devices, the majority of users stated a preference for using one hand, regardless of task or device owned. Users of *TS-with-qwerty* devices preferred two hands for text messaging, email composition, and text entry. Based on these data, it is consistent that 66 percent of participants stated they would prefer to use one hand for the majority of device interaction, versus nine percent who would prefer two hands for all interaction. Twenty-three percent did not have a preference and six users did not respond.

Survey Summary

Considering current usage patterns only, there is no obvious winner between one and two handed device use. Excluding phone calls, the number of activities for which a majority of respondents use one (seven) vs. two hands (six) is nearly balanced. However, device type certainly influences user behavior; users of keypad-only devices nearly always use one hand, while users of touchscreen devices more often favor two hands, especially for tasks involving text entry. But user justifications for hand choice indicate that the hardware/software interface is to blame for much two-handed use occurring today. Most use one hand if at all possible and only use two hands when the interface makes a task impossible to do otherwise. Other than gaming, tasks involving text entry are the only ones for which users may be willing to use two hands, especially when the device used provides an integrated Qwerty keyboard. It seems, therefore, that the efficiency gained by using two hands for such tasks is often worth the dedication of physical resources, which is also true of the immersive gaming experience.

While most users can imagine the ideal of single-handed text entry, enabling single-handed input may not be enough—*throughput* is also important. Ultimately, it is clear that interface designers of all device types should make one-handed usability a priority, and strive to bridge the gap between current and desired usage patterns.

THUMB MOVEMENT STUDY

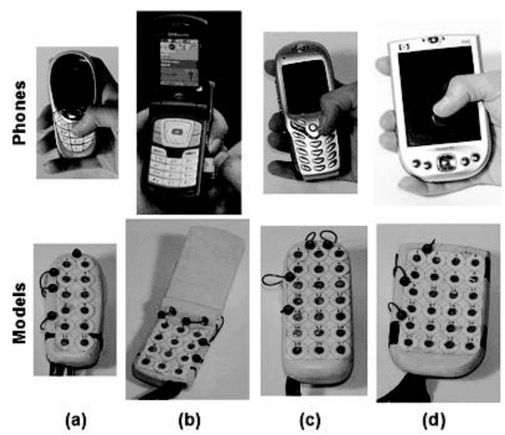
The third component of our exploration was an examination of thumb movement in the context of mobile device interaction. As input technologies and device forms come and go, biomechanical limitations of the thumb will remain. Although the thumb is a highly versatile appendage with an impressive range of motion, it is most adapted for grasping tasks, playing opposite the other four fingers (Bourbonnais, Forget, Carrier et al., 1993). Hence thumb interaction on the surface of today's mobile devices introduces novel movement and exertion requirements for the thumb—repetitive pressing tasks issued on a plane parallel to the palm. We believe a fundamental understanding of thumb capabilities when holding a device can help guide the placement of interaction targets for both hardware and software interfaces designed for one-handed use. Although we can make reasonable guesses about thumb capabilities, empirical evidence is a better guide. Since no strictly relevant studies have yet been conducted, we developed a study to help us understand how device form and task influences thumb mobility.

Since thumb tapping is the predominant means of interaction for keypad-based devices, and has also proven promising for one-handed touchscreen use (Karlson et al., 2005), we focused our investigation on surface tapping tasks. We hypothesized that the difficulty of a tapping task would depend on device size, movement direction, and surface location of the interaction. We captured the impact of these factors on user performance by using movement speed as a proxy for task difficulty, under the assumption that harder tasks would be performed more slowly than easier tasks.

Equipment

Device models. For real devices, design elements such as buttons and screens communicate to the user the "valid" input areas of the device. We instead wanted outcomes of task performance to *suggest* appropriate surface areas for thumb interaction. We identified four common hand-held devices to represent the range of sizes and shapes found in the market today: (1) a Siemens S56 candy bar phone measuring 4.0 x 1.7 x 0.6 in (10.2 x 4.3 x 1.5 cm); (2) a Samsung SCH-i600 flip phone measuring 3.5 x 2.1 x 0.9 in (9 x 5.4 x 2.3 cm); (3) an iMate smartphone measuring 4 x 2.0 x 0.9 in (10.2 x 5.1 x 2.3 cm) and (4) an HP iPAQ h4155 Pocket PC measuring 4.5 x 2.8 x 0.5 in (11.4 x 7.1 x 1.3 cm). These devices are shown in the top row of Figure 4. We refer to these as simply SMALL, FLIP, LARGE, and PDA. To remove the bias inherent in existing devices, we created a 3D model of each device, removing all superficial design features. The models were developed using Z Corp.'s (http://www.zcorp.com/) ZPrinter 310 3D rapid prototyping system. Device models were hollow, but we reintroduced weight to provide a realistic feel. Once "printed"

Figure 4. Thumb movement study: Devices we chose to represent a range of sizes and forms (top row) together with their study-ready models (bottom row): (a) SMALL, (b) FLIP, (c) LARGE, and (d) PDA.



and cured, the models were sanded and sealed to achieve a smooth finish.

Target design. A grid of circular targets 1.5 cm in diameter was affixed to the surface of each device. Circles were used for targets so that the sizes would not vary with direction of movement (MacKenzie & Buxton, 1992). The target size was selected to be large enough for the average-sized thumb, while also providing adequate surface coverage for each device. The grid dimensions for each device were: SMALL (2x5), FLIP (3x4), LARGE (3x7) and PDA (4x6), as shown in the bottom row of Figure 4.

Measurement. A typical measurement strategy for tapping tasks would involve a surface-based sensor to detect finger contact. Unfortunately, due to the number and variety of device sizes investigated, no technical solution was found to be as versatile, accurate or affordable as required. Instead we used Northern Digital Inc's OPTOTRAK 3020 motion analysis system designed for fine-grained tracking of motor movement. The OPTOTRAK uses 3 cameras to determine the precise 3D coordinates of infrared emitting diodes (IREDs). Three planar IREDs attached to the surface of each device defined a local coordinate system, and a fourth IRED provided redundancy (see Figure 4, bottom row). The spatial positions of two markers affixed to each participant's right thumb were then translated with respect to the coordinate system of the device to establish relative movement trajectories. Diode positions were sampled at 100Hz, and data were post processed to derive taps from thumb minima.

Software. Data collection and experiment software was run on a Gateway 2000 Pentium II with 256 MB of RAM running Windows 98.

Participants

Twenty participants were recruited via fliers posted in our department of computer science, with the only restriction that participants be right-handed. Participants (15 male) ranged in age from 18 to 35 years with a median age of 25 years. Participants received $20 for their time.

Design

For each target on each device (SMALL, FLIP, LARGE, and PDA), users performed all combinations of *distance* (1 or 2 circles) x *direction* (↕, ↔, ⤢ , ⤡) tasks that could be supported by the geometry of the device. For example, SMALL could not accommodate trials of distance 2 circles in the directions (↔, ⤢, ⤡). Note that the grid layout results in *actual* distances that differ between *orthogonal* trials (↕, ↔) and *diagonal* trials (⤢, ⤡), which we consider explicitly in our analysis. For LARGE and PDA, trials of distance 4 circles were included as the geometry permitted. Finally each device included a ⤢ and ⤡ trial to opposite corners of the target grid. For each device, a small number of trials (1 for SMALL, LARGE and PDA, 3 for FLIP), selected at random, were repeated so as to make the total trial count divisible by four. The resulting number of trials for each device were: SMALL (32), FLIP (48), LARGE (108), and PDA (128). Since the larger devices had more surface targets to test, they required more trials.

Tasks

Users performed reciprocal tapping tasks in blocks as follows. For SMALL and FLIP, trials were divided equally into two blocks. For the LARGE and PDA, trials were divided equally into four blocks. Trials were assigned to blocks to achieve roughly equal numbers of *distance* x *direction* trials, distributed evenly over the device. Trials were announced by audio recording so that users could focus attention fully on the device. Users were presented with the name of two targets by number. For example, a voice recording would say "1 and 3." After one second, a voice-recorded "start" was played. Users tapped as quickly as possible between the two targets, and after five seconds, a "stop" was played. After a 1.5 second delay the next trial began. Trials continued in succession to the end of the block, at which point the user was allowed to rest as desired, with no user resting more than two minutes. Device and block orders were assigned to subjects using a Latin Square, but the presentation of within-block trials was randomized for each user.

Procedure

Each session began with a brief description of the tasks to be performed and the equipment involved. Two IRED markers were then attached to the right thumb with two-sided tape. One diode was placed on the leftmost edge of the thumb nail, and a second on the left side of the thumb. The orthogonal placement was intended to maximize visibility of at least one of the diodes to the cameras at all times. The two marker wires were tethered loosely to the participant's right wrist with medical tape.

The participant was seated in an armless chair, with the device held in the right hand, and the OPTOTRAK cameras positioned over the right shoulder. At this point the participant was given more detailed instruction about the tasks, and informed of the error conditions that might occur during the study: if at any point fewer than three of the device-affixed IREDs or none of the thumb IREDs were visible to the cameras, an out-of-sight error sound would be emitted, at which point he or she should continue the trial as naturally as possible while attempting to make adjustments to improve diode visibility. Next, the participant was given the first device and performed a practice session of 24 trials, selected to represent a variety of distances, directions, and surface locations. During the practice trials, the administrator intentionally occluded the diodes to give the participant familiarity with the out-of-sight error sound and proper remedies. After completion of the practice trials and indication that the participant was ready, the study proper was begun.

During trials, participants were allowed to hold the devices in whatever manner supported their best performance. Since the instructions were presented audibly and with a short pause before the trial began, users could prepare their grip if desired. We chose not to control for grip in our study under the reasoning that it resembled real world settings, in which users have the freedom to adjust their grips to best suit the environment and task.

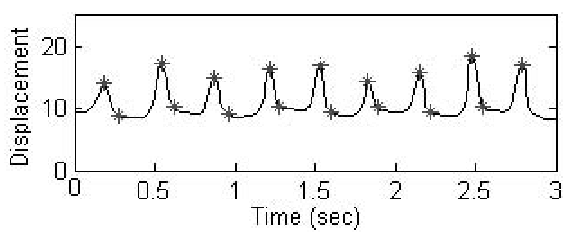
After all trials for a device were completed, users were allowed to rest while the next device was readied, typically three to five minutes; together with the rest period users were offered between trial blocks and counterbalanced device order, we hoped to minimize as much as possible the impact of fatigue on the results. After completing all trials for the last device, the participant completed a questionnaire, recording demographics and subjective ratings. Total session time was two hours, approximately an hour of which was devoted to data collection.

Measures

Raw 3D thumb movement data for each five second trial were truncated to the middle three seconds to eliminate artifacts resulting from initiation lag and anticipated trial completion, phenomena routinely observed by the administrator. In a post processing phase, taps were identified within the remaining three second interval and a single average tap time was computed from the difference in time between the onset of the first tap to the onset of the last tap, divided by one fewer than the total number of taps detected. In a post experiment questionnaire, participants assigned an overall rating of difficulty to each device (1-7, where 1 = easy, 7 = difficult), and indicated the device regions that were both easiest and hardest to interact with.

Data post processing. Since the 3D thumb position (x,y,z) was recorded relative to the device surface, the z-value represented the thumb height above the device. While one might assume that taps were those thumb positions for which the z-distance was 0, the IREDs were mounted on participants' thumbnails, and so never actually

Figure 5. Example MATLAB output of the thumb's distance from the surface of the device. Stars depict the auto-detected peaks and valleys.



reached the surface of the device. Taps were instead defined as points when both the *z*-value and change in *z*-value (velocity) were minimal. For example, plotting *z*-values over time reveals a wave pattern whose valleys indicate taps (Figure 5).

Raw data was first preprocessed to extract the middle three seconds of each trial as well as to select the thumb diode with the most complete data set (e.g., the fewest number of missing frames, or if equal, the one with the most compact windows of missing frames). Linear interpolation was performed on missing frames if the gap was less than 100 ms. Missing frames included those lost due to out-of-sight errors, as well as occasional frames dropped by the collection hardware.

The data was then analyzed by the PICKEXTR MATLAB function to identify extrema in a signal. This function is provided with the RelPhase.Box Matlab toolbox for relative phase analysis of oscillatory systems (Dijkstra, Giese, & Schöner, 1997). The accuracy of the tap classifier was verified by inspecting a visual representation (Figure 5) of each trial. When required, corrections were made as follows: (1) valid endpoints were preserved, (2) if intermediate taps were missing, they were added, (3) if intermediate taps were incorrect, they were recoded by hand, and (4) if endpoints were invalid, the entire signal was coded by hand. Since average tap time was calculated as the number, not placement, of intervening taps, this method minimized as much as possible the bias of human annotation. Of the trials included for statistical analysis, 1.3 percent were discarded because they could not be encoded by machine or human, or had less than 1.5 seconds of encodable signal.

Results

The goal of our analysis was to understand whether user performance was influenced by device size, interaction location, and movement direction. To allow for comparison among the devices, we limited the analysis to trials with distances of 1 or 2 circles since the geometries of all but the smallest device (SMALL) supported these trials in all four movement directions. To address the fact that actual movement distance differed between orthogonal and diagonal trials, we analyzed these groups separately. For all analyses, Huynh-Felt corrections were used when the sphericity assumption was violated, and Bonferroni corrections were used for post hoc comparisons.

Movement Direction. A 2 (*distance*) x 2 (*direction*) repeated measures analysis of variance (RM-ANOVA) was performed on mean task time data for both orthogonal trials (distances: 1, 2; directions: ↕, ↔) and diagonal trials (distances: 1.4, 2.8; directions: ⤢, ⤡) for the three largest devices. Since SMALL did not support distance 2 trials in all four directions, a one-way RM-ANOVA was performed on mean task time for trials of distance 1 and 1.4.

SMALL: A main effect of direction was observed for diagonal trials ($F(1,19) = 65.1$, $p < .001$). Post hoc analyses showed that trials in the ⤢ direction were performed significantly faster than those in the ⤡ direction (0.26 v. 0.28 ms, $p < .001$).

FLIP, *LARGE*, *and PDA*: Results were similar across the analyses of the three largest devices. Unsurprisingly, a main effect of distance was observed for both orthogonal and diagonal trials, with shorter trials significantly faster than longer trials. There were no further effects of direction or interaction between direction and distance for orthogonal trials (↕, ↔). However, for diagonal trials, a main effect of direction was observed, with trials in the ⤢ direction significantly faster than those in the in ⤡ direction for all devices. In addition, a *distance* x *direction* interaction showed performance differences between the diagonal trials were more pronounced for longer trials than shorter trials (Table 1).

Device Size

To determine if device size impacted comparable tasks across devices, we analyzed all trials performed in the lower right 3x4 region of the three largest devices using a 3 (*devices*) x 43 (*trials*) RM-ANOVA. While a main effect of trial was observed, this was expected, as trials of every distance and direction were included for analysis. Yet no effects of *device* or *device* x *trial* were found.

Table 1. Mean time for movement direction and distance x direction for FLIP, LARGE, and PDA

	Direction (⤢ v. ⤡)	$F_{1,19}$	p	Dist. x Dir. (⤢ v. ⤡)	$F_{1,19}$	p
FLIP	.31 v. .35 ms	50.5	<.001	(1.4) .27 v. .30 ms (2.8) .34 v. .41 ms	14.6 34.5	<.001 <.001
LARGE	.31 v. .36 ms	46.1	<.001	(1.4) .27 v. .28 ms (2.8) .35 v. .42 ms	28.0 44.2	<.001 <.001
PDA	.32 v. .36 ms	46.5	<.001	(1.4) .28 v. .30 ms (2.8) .36 v. .43 ms	23.0 38.0	<.001 <.001

Target Location

To determine if target location affected performance, we analyzed task time for the shortest tasks for each device. We chose short tasks because they provide high granularity for discriminating among device locations. Since direction was shown to affect task time for diagonal trials, only orthogonal tasks could be considered. For each device, a one-way RM-ANOVA was performed on mean trial time, with the number of trials varying by device.

A main effect of target location was observed for SMALL (F (8.6, 163.3) = 2.1, p = .032), FLIP (F (11.5, 218.4) = 3.5, p < .001) and PDA (F (9.8, 188.1) = 3.9, p < .001), but not for LARGE. However, in post hoc analyses, only PDA had a reasonable number of trials that differed significantly from one another. Since it is difficult to draw helpful conclusions from specific pairs of trials, we explored two aggregation techniques.

Subject-derived regions. Based on subjective opinion of which regions were easiest to reach for each device, we divided tasks into three groups (E)asy, (M)edium, and (H)ard. Tasks for SMALL and FLIP were assigned to only E and M groups. A one-way RM-ANOVA on mean group task time was performed for each device. A main effect of group was found for FLIP (F (5.5, 105.1) = 11.3, p < .001), LARGE (F (3.1, 58.7) = 8.4, p < .001), and PDA (F (4.8, 91.0) = 22.0, p < .001). Post hoc analyses showed all groups differed significantly from each other for FLIP and PDA. For LARGE, E and M were significantly faster than H, but were indistinguishable otherwise, so we collapsed them to E (Table 2, third column).

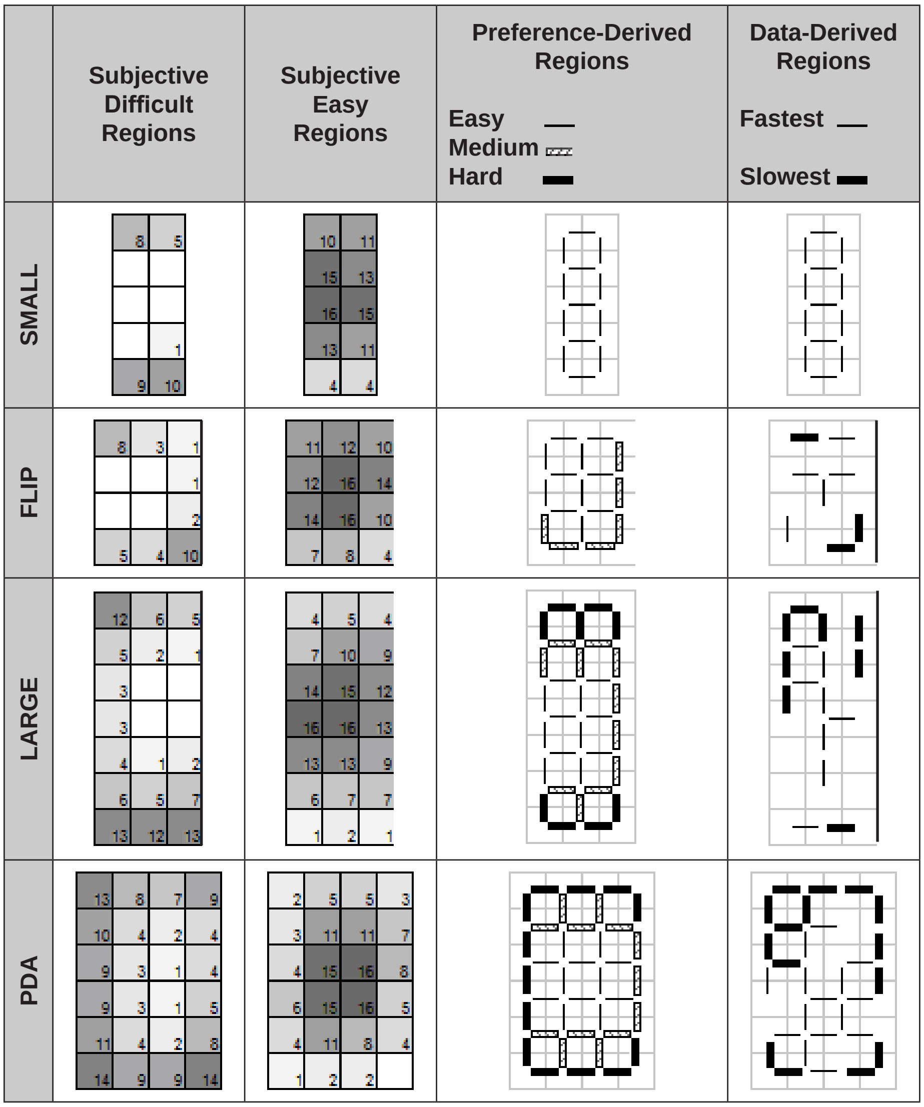
Data-derived regions. For each device we ordered tasks by mean tap time, and then segmented them into seven groups. If the number of trials was not divisible by seven, the remainder trials were included in the middle group. A one-way RM-ANOVA on mean group task time was performed for each device. A main effect of group was found for FLIP (F (5.5, 105.1) = 11.3, p < .001), LARGE (F (3.1, 58.7) = 8.4, p < .001), and PDA (F (4.8, 91.0) = 22.0, p < .001). From these results, groups were labeled *fastest* and *slowest* such that all groups in *fastest* were significantly faster than all groups in *slowest*, according to post hoc analyses. Trials in these groups are shown visually in the rightmost column of Table 2. Mean task time for *fastest* v. *slowest* trials for each device were FLIP (0.26 v. 0.28 ms), LARGE (0.25 v. 0.28 ms), and PDA (0.26 v. 0.29 ms).

Subjective Preferences

After completing all trials, users were presented with diagrams of each device similar to those in the first two columns of Table 2 and asked to identify the targets they found most easy and most difficult to interact with. Aggregating results across users yielded a preference "map" for the least and most accessible targets of each device (columns 1 and 2 of Table 2), with darker regions indicating more agreement among participants. We see that for each device the two representations are roughly inverses of one another.

In addition to region marking, we asked users to rate the overall difficulty of managing each device with one hand on a 7-point scale (7 = most

Table 2. Preference and movement time maps for each device. Depth of color in columns 1 and 2 indicate stronger user agreement.



comfortable). Average ratings from most to least comfortable were as follows: SMALL (6.4), FLIP (5.4), LARGE (4.1) and PDA (3.0).

Thumb Movement Summary

The findings from our analysis of thumb movement suggest the following guidelines. First, thumb movement in the ↘ direction is difficult for right-handed users regardless of device size. Presumably the difficulty arises from the considerable flexion required to perform these types of tasks. Under this reasoning, the opposite movement ↗ would be difficult for left-handed users, so conservative designs should constrain repetitive movement to ↕ and ↔ directions to accomodate all users, and especially for repetitive tasks such as text or data entry.

Second, device region affects both task performance and perceived difficulty. Not only did the slowest trials correspond to those regions users found most difficult, but fastest trials also matched those regions users found most easy (Table 2). In general, regions within reach of the thumb were fastest and most comfortable, favoring those toward the midline of the device—a "sweet spot" that required movement primarily from the base of the thumb. The lower right corners of the devices present an exception in that they are biomechanically awkward to reach because they are "too close" rather than "too far."

Because the absolute time differences between the fastest and slowest regions of the devices were quite small (at most 30 ms), we do not think performance speed is the main concern in forming design recommendations from these data. Rather, it is the fact that the speed differences between the regions were statistically significant (7%-12% slowdown between the fastest and slowest regions) that suggests a mechanical and/or physical encumbrance was to blame. The data, therefore, are concerning primarily from an ergonomics perspective. In fact, we believe that the slowdowns we found should be thought of as optimistic, since they capture only localized movement and required substantial changes in user grip between tasks; subjective opinion, user observations and practical experience indicate that designers should be cautioned against using the entire device surface for thumb interaction, especially for larger devices. We instead recommend placing interaction objects centrally to accommodate both left and right handed users, or offering configurable displays. Since hand size and thumb length will differ by individual, designs should strive to support a range of users.

Finally, the result that users performed trials in the lower right 3x4 sub-grid of the three largest devices equally well suggests that large devices do not inherently impede thumb movement. Rather, larger devices simply have more areas that are out of thumb reach, and so have more regions that are inappropriate for object placement in one-handed designs. Together with user opinion that larger devices were more difficult to manage suggests that the current trend toward smaller device forms benefits one handed device control.

FUTURE TRENDS

Market forces are driving the sizes of mobile devices down and at the same time driving feature sets up. These concurrent trends *alone* compromise the ease and expressiveness of information interaction and presentation, as shrinking keypads and screens are being used to access ever increasing data sets. Yet at the same time, mobile users often are carrying personal effects, opening doors, holding handrails, or otherwise needing the use of a hand to manage the environment, which then leaves only a single hand available for device operation. Indeed, results from our field study and survey of mobile device use have confirmed that one-handed device operation is widespread—not limited to a niche user segment—and that users would prefer to use one hand more often than current designs allow.

Commercial trends for supporting single handed device operation have focused on thumb operation of touchscreen based devices. Several thumb-based virtual keypads are available from third party vendors, such as the Phraze-It® keypad from Prevalent Devices, www.prevalentdevices.com. One of the few handheld *systems* dedicated to single-handed operation is the touchscreen-based N1m phone by Neonode, www.neonode.com, which supports application navigation and interaction using only thumb taps and sweeps. However, the N1m's primary use is as a phone, camera and media player, rather than a personal data manager. As such, the N1m is not designed to support rich graphical interfaces or data interactions.

As the mobile user base expands, so do device storage capacities and wireless services. Not only are mobile devices accumulating more resident data, but they are increasingly used as front-end interfaces to external data sets. Given the broad range of data sets and tasks users expect of today's devices, single-handed device support will need to generalize beyond specific applications and technologies seen today. Our work seeks to understand some of the basic human factors involved in single handed device use, in order that this knowledge can be applied generally to the variety of tasks and device forms users demand for portable computing.

CONCLUSION

In an effort to understand the one-handed interaction needs of mobile device users, we looked at a broad range of device use. Our field study showed that for at least one class of user (travelers), mobile phones are most often used with one hand, and that this behavior seems to correlate with activity, such as walking or holding items in the other hand. Our survey revealed that the vast majority of users want to use one hand for interacting with mobile devices, but that current interfaces, especially for touchscreens, are not designed to support dedicated single handed use. Finally, an empirical evaluation of thumb interaction on varying-sized devices suggests that (1) mid-device regions are easiest to access; (2) the position of a target with respect to the thumb impacts performance more than device size, and finally (3) ⤡ movement is difficult for right-handed users and degrades with movement distance.

REFERENCES

Apple. (2006). *iPod*. Retrieved January 16, 2007, from http://www.apple.com/ipod/ipod.html.

Barmakian, J. T. (1992). Anatomy of the joints of the thumb. *Hand Clinics, 8*(4), 681-691.

Barnard, L., Yi, J. S., Jacko, J. A., & Sears, A. (2005). An empirical comparison of use-in-motion evaluation scenarios for mobile computing. *International Journal of Human-Computer Studies, 62*, 487-520.

Bourbonnais, D., Forget, R., Carrier, L., & Lepage, Y. (1993). Multidirectional analysis of maximal voluntary contractions of the thumb. *Journal of Hand Therapy, 6*(4), 313-318.

Department of Health and Human Services. (2005). 45 CFR Part 46, Protection of Human Subjects. In (Section 46.101(b)). Retrieved January 16, 2007 from http://www.hhs.gov/ohrp/humansubjects/guidance/45cfr46.htm

Dijkstra, T. M., Giese, M. A., & Schöner, G. (1997). RelPhase.box: A Matlab toolbox for relative phase analysis of oscillatory systems. Retrieved January 17, 2007, from http://www.psy.ohio-state.edu/visionlab/dijkstra/Software/RelPhase.sit

Dong, L., Watters, C., & Duffy, J. (2005). Comparing two one-handed access methods on a PDA. *International Conference on Human Computer Interaction with Mobile Devices and Services*, (pp. 235-238). ACM Press.

Hinckley, K., Pierce, J., Sinclair, M., & Horvitz, E. (2000). Sensing techniques for mobile interaction. *ACM Symposium on User Interface Software and Technology*, (pp. 91-100). ACM Press.

Hirotaka, N. (2003). Reassessing current cell phone designs: using thumb input effectively. *Extended Abstracts of the ACM Conference on Human Factors in Computing*, (pp. 938-939). ACM Press.

Karlson, A. K., Bederson, B. B., & SanGiovanni, J. (2005). AppLens and LaunchTile: two designs for one-handed thumb use on small devices. *ACM Conference on Human Factors in Computing*, (pp. 201-210). ACM Press.

Kristoffersen, S., & Ljungberg, F. (1999). Making place to make it work: empirical exploration of HCI for mobile CSCW. *ACM Conference on Supporting Group Work*, (pp. 276-285). ACM Press.

Kuo, L., Cooley, W., Kaufman, K., Su, F., & An, K. (2004). A kinematic method to calculate the workspace of the TMC joint. In *Proceedings of the Institute of Mechanical Engineering H, 218*(2), 143-149.

Li, Z.-M., & Harkness, D. A. (2004). Circumferential force production of the thumb. *Medical Engineering and Physics, 26*, 663-670.

Lyons, K., Starner, T., Plaisted, D., Fusia, J., Lyons, A., Drew, A. et al. (2004). Twiddler typing: one-handed chording text entry for mobile phones. *ACM Conference on Human Factors in Computing*, (pp. 671-678). ACM Press.

MacKenzie, I. S., & Buxton, W. (1992). Extending Fitts' Law to two-dimensional tasks. *ACM*

Conference on Human Factors in Computing, (pp. 219-226). ACM Press.

Milanesi, C., Liang, A., Vergne, H.J.D.L., Mitsuyama, N., & Nguyen, T.H. (2007). *Data insight: Market share for mobile devices, 1*(7). Gartner, Inc.

Oulasvirta, A., Tamminen, S., Roto, V., & Kuorelahti, J. (2005). Interaction in 4-second bursts: the fragmented nature of attentional resources in mobile HCI. *ACM Conference on Human Factors in Computing*, (pp. 919-928). ACM Press.

Pascoe, J., Ryan, N., & Mores, D. (2000). Using while moving: HCI issues in fieldwork environments. *Transactions on Computer-Human Interaction, 7*(3), 417-437.

Pirhonen, P., Brewster, S. A., & Holguin, C. (2002). Gestural and audio metaphors as a means of control in mobile devices. *ACM Conference on Human Factors in Computing*, (pp. 291-298). ACM Press.

Rekimoto, J. (1996). Tilting operations for small screen interfaces. *ACM Symposium on User Interface Software and Technology,* (pp. 167-168). ACM Press.

Wigdor, D., & Balakrishnan, R. (2003). TiltText: using tilt for text input to mobile phones. *ACM Symposium on User Interface Software and Technology,* (pp. 81-90). ACM Press.

KEY TERMS

Biomechanics: The study of muscular mechanics and activity.

Candybar Phone: A traditional-style mobile phone with a rigid rectangular form, typically about 3 times longer than wide.

Ergonomics: Often used synonymously with “human factors,” ergonomics is the scientific discipline concerned with the understanding of interactions among humans and other elements of a system, and the profession that applies theory, principles, data and methods to design in order to optimize human well-being and overall system performance (definition adopted by the International Ergonomics Association in August 2000).

Flip Phone: A popular form factor for a mobile phone that structures the phone into two halves that are hinged like a clamshell: the numeric keypad is placed on the lower half, and the display is placed on the upper half. In the closed position the screen and the keypad are protected from inadvertent damage or activation, while in the open position users can angle the screen for optimal visibility.

Mobile or Cellular Phone: A portable electronic device which, at a minimum, supports long-range telecommunications. Today’s mobile phones support a much broader range of functions for personal data and data management activities, including data exchange with a personal desktop computer, music play, and photo capturing.

PDA: Personal digital assistant. The original handheld device designed expressly for supporting personal data management on the go. PDAs typically have touch sensitive screens and are operated with a stylus. Recent models may also include a miniaturized Qwerty keyboard for text entry.

Chapter VII
User Acceptance of Mobile Services

Eija Kaasinen
VTT Technical Research Centre of Finland, Finland

ABSTRACT

Personal mobile devices are increasingly being used as platforms for interactive services. User acceptance of mobile services is not just based on usability but includes also other interrelated issues. Ease of use is important, but the services should also provide clear value to the user and they should be trustworthy and easy to adopt. These user acceptance factors form the core of the Technology Acceptance Model for Mobile Services introduced in this chapter. The model has been set up based on field trials of several mobile services with altogether more than 200 test users. The model can be used as a design and evaluation framework when designing new mobile services.

INTRODUCTION

Research on mobile services has thus far mainly concentrated on the usability of alternative user interface implementations. Small mobile devices pose significant usability challenges and the usability of the services is still worth studying. However, more attention should be paid to user acceptance of the planned services. The reason for many commercial failures can be traced back to the wrongly assessed value of the services to the users (Kaasinen, 2005b).

User evaluations of mobile services often have to be taken into the field as the service would not function properly otherwise, or it would not make sense to evaluate it in laboratory conditions. This would be the case, for instance, with GPS systems and route guidance systems. In long-term field trials with users, it is possible to gather feedback on the adoption of the service in the users' everyday lives. Such studies gather usage data beyond mere usability and pre-defined test tasks (Figure 1). Field trials help in studying which features the users start using, how they use them and how often, and which factors affect user acceptance of the service.

Figure 1. Taking user evaluations from the laboratory to the field makes it possible to evaluate user acceptance on new services

Business and marketing research already have approaches whereby new technology is studied on a wider scale. The Technology Acceptance Model by Davis (1989) defines a framework to study user acceptance of a new technology based on perceived utility and perceived ease of use. Each user perceives the characteristics of the technology in his or her own way, based for instance on his or her personal characteristics, his or her attitudes, his or her previous experiences and his or her social environment. The Technology Acceptance Model has been evolved and applied widely, but mainly in the context of introducing ready-made products rather than in designing new technologies.

In this chapter an extension to the Technology Acceptance Model will be introduced. The model is based on a series of field trials and other evaluation activities with different mobile Internet and personal navigation services and over 200 test users (Kaasinen, 2005b). The Technology Acceptance Model for Mobile Services constitutes a framework for the design and evaluation of mobile services.

BACKGROUND

Technology acceptance models aim at studying how individual perceptions affect the intentions to use information technology as well as actual usage (Figure 2).

In 1989, Fred Davis presented the initial technology acceptance model (TAM) to explain the determinants of user acceptance of a wide range of end-user computing technologies (Davis 1989). The model is based on the Theory of Reasoned Action by Ajzen and Fishbein (1980). TAM points out that perceived ease of use and perceived usefulness affect the intention to use. Davis (1989) defines perceived ease of use as "*the degree to which a person believes that using a particular system would be free from effort*" and perceived usefulness as "*the degree to which a person believes that using a particular system would enhance his or her job performance*." Perceived ease of use also affects the perceived usefulness (Figure 3). The intention to use affects the real usage behavior. TAM was designed to study information systems at work to predict if the users will actually take a certain system into use in their jobs. The model provides a tool to study the impact of external variables on internal beliefs, attitudes and intentions.

TAM deals with perceptions; it is not based on observing real usage but on users reporting their conceptions. The instruments used in connection with TAM are surveys, where the questions are constructed in such a way that they reflect the different aspects of TAM. The survey questions related to usefulness can be, for instance: "Using this system improves the quality of the work I do" or "Using this system saves my time." The survey questions related to ease of use can be, for instance:

Figure 2. The basic concept underlying technology acceptance models (Venkatesh et al., 2003)

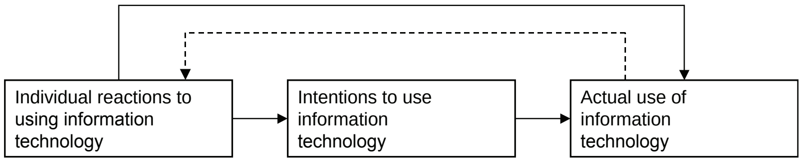
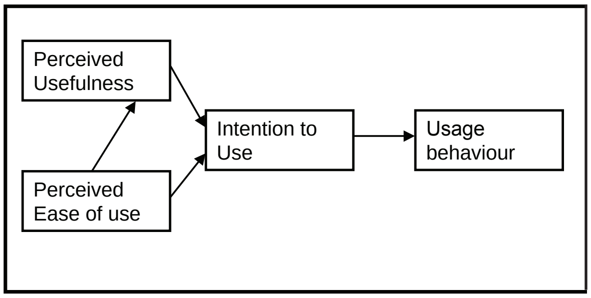


Figure 3. Technology acceptance model (Davis, 1989)



"The system often behaves in unexpected ways" or "It is easy for me to remember how to perform tasks using this system."

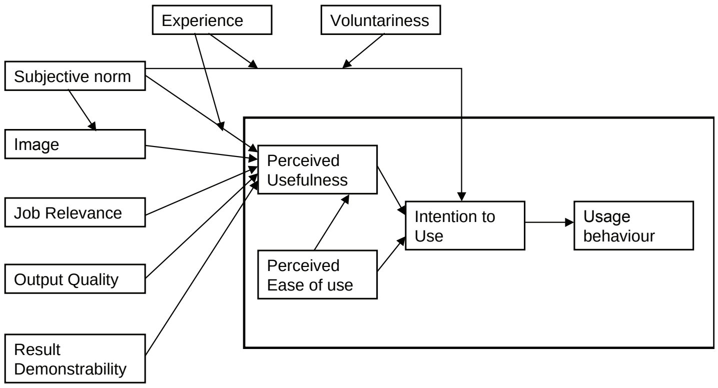
TAM has been tested and extended by many researchers, including Davis himself. Venkatesh and Davis (2000) have enhanced the model to TAM2 (Figure 4), which provides a detailed account of the key forces underlying judgments of perceived usefulness, explaining up to 60 percent of the variance in this driver of usage intentions. TAM2 showed that both social influence processes (subjective norm, voluntariness and image) and cognitive instrumental processes (job relevance, output quality, result demonstrability, and perceived ease of use) significantly influenced user acceptance.

Mathieson, Peacock and Chin, (2001) have extended TAM by analyzing the influence of perceived user resources. They claim that there may be many situations in which an individual wants to use an information system, but is prevented by lack of time, money, expertise and so on (Mathieson et al., 2001) classify resource-related attributes into four categories: user attributes, support from others, system attributes and general control-related attributes that concern an individual's overall beliefs about his or her control over system use. In their extended model, external variables affect perceived resources that further affect perceived ease of use and the intention to use.

TAM was originally developed for studying technology at work, but it has often been used to study user acceptance of Internet services as well (Barnes & Huff, 2003; Chen, Gillenson & Sherell, 2004; Gefen, 2000; Gefen & Devine, 2001; Gefen, Karahanna & Straub, 2003). Gefen et al. (2003) have studied TAM in connection with e-commerce. They have extended TAM for this application area and propose that trust should be included in the research model to predict the purchase intentions of online customers.

The Technology Acceptance Model constitutes a solid framework to identify issues that may affect user acceptance of technical solutions. Davis and Venkatesh (2004) proved that the model can be enhanced from the original purpose of

Figure 4. Enhanced technology acceptance model (TAM2) by Venkatesh and Davis (2000)



studying user acceptance of existing products to study planned product concepts, for example, in the form of mock-ups. This indicates that TAM could also be used in connection with technology development projects and processes to assess the usefulness of proposed solutions.

APPLICABILITY OF EARLIER APPROACHES FOR MOBILE SERVICES

The focus of traditional usability studies is on specified users performing specified tasks in specified contexts of use (ISO13407, 1999). In field trials the users can use prototype services as part of their everyday life. The research framework can then be enhanced to identify the actual tasks that users want to perform and the actual contexts of use. Technology acceptance models provide a framework for such studies.

Mobile services targeted at consumers have several specific characteristics that may mean that their user acceptance cannot be studied using the same models as with information systems in the workplace. When dealing with consumer services, individuals make voluntary adoption decisions and thus the acceptance includes assessing the benefits provided compared with either competing solutions or the non-acquisition of the service in question. As pointed out by Funk (2004), mobile services are disruptive technology that may find their innovation adopters elsewhere than expected, as highlighted by the experiences with the Japanese i-mode. Focusing too early on only limited user groups may miss possible early adopters. With the Japanese i-mode, other services were boosted through e-mail and personal home pages (Funk, 2004). This suggests that the focus of user acceptance studies of mobile services should be extended to interrelated innovations, as proposed by Rogers (1995).

Perceived usefulness included in TAM may not indicate an adequate purchase intention in a market situation. Product value has been proposed as a wider design target both in software engineering and HCI approaches. A value-centered software engineering approach was proposed by Boehm (2003) to define more clearly what the design process is targeted at, and identifying the values that different stakeholders—including end-users—expect of the product. Although not using the actual term "value," Norman (1998) emphasizes the importance of identifying big phenomena related to user needs and communicating them early on to the design. Cockton (2004b) points out that in value-centered HCI existing HCI research components, design guidance, quality in use and fit to context need to be reshaped to subordinate them to the delivery of product value to end-users and other stakeholders.

Mobile services are increasingly handling personal information of the user, for instance due to the personalization and context-awareness of the services. The functionalities of the increasingly complex systems are not always easy for the users to comprehend. Context-aware services may include uncertainty factors that the users should be able to assess. Mobile service networks are getting quite complex and the users may not know with whom they are transacting. Technical infrastructures as well as the rapidly developed services are prone to errors. All these issues raise trust as a user acceptance factor, similar to TAM applied in e-commerce (Chen et al., 2004; Gefen et al., 2003). Trust has been proposed as an additional acceptance criterion for mobile services by Kindberg, Stellen and Geelhoed, (2004) and Barnes and Huff (2003). Trust has also been included in studies of personalization in mobile services (Billsus et al., 2002) and studies of context-aware services (Antifakos, Schwaninger & Schiele, 2004).

Ease of adoption is included in the studies by Sarker and Wells (2003) and Barnes and Huff (2003). Sarker and Wells (2003) propose a totally new acceptance model that is based on user adoption. Barnes and Huff (2003) cover adoption in their model within the wider themes of compatibility and trialability. *Perceived user resources* in the extension of TAM by Mathieson et al. (2001) and *Facilitating conditions*, in the Unified Theory of Acceptance and Use (Venkatesh et al., 2003) also include elements related to ease of adoption.

In the following, the technology acceptance model for mobile services (Kaasinen, 2005b) is described in detail. The model aims at taking into account the aforementioned special characteristics of mobile consumer services, and previous studies on user acceptance described in this chapter. The model can be utilized when designing new services and assessing them to ensure that key user acceptance factors are considered in the design.

TECHNOLOGY ACCEPTANCE MODEL FOR MOBILE SERVICES (TAMM)

The technology acceptance model for mobile services (TAMM) was constituted based on a series of field trials and other user evaluation activities involving over 200 users. The studies were carried out as parts of technology development projects in 1999-2002 by project usability teams comprising altogether 13 researchers from VTT and three researchers from other research organizations. The focus of the studies was in particular on mobile Internet services and location-based services targeted at consumers (Kaasinen, 2005b). Mobile Internet studies were carried out in connection with the development of mobile browsers and the first WAP (wireless application protocol) services for mobile phones. In addition to commercial services, the test users could access many Web services because our project developed a Web-WAP conversion proxy server. Based on identified user needs, our research team also developed specific WAP services, for instance, for group communication. The services were evaluated in long-term field trials with users. The studies of location-based services were carried out within a horizontal usability support project, part of the Personal Navigation (NAVI) research and development program in Finland. The aim of the program was to facilitate co-operation between different actors who were developing personal navigation products and services. Our research group supported individual projects in usability and ethical issues and, beyond this, identified general guidelines for acceptable personal navigation services. We studied user attitudes and preliminary acceptance by evaluating different service scenarios in focus groups. In addition we evaluated some of the first commercial location-based services and carried out user evaluation activities in co-operation with the NAVI projects that were developing location-based services. Table 1 gives an overview of the user evaluation activities that the technology acceptance model for mobile services is based on.

The original technology acceptance model was chosen as the starting point for the new model because it provided a framework for connecting field study findings of ease of use and usefulness. The user acceptance framework is especially suitable for field trials where the focus is to study how different users start using the mobile services

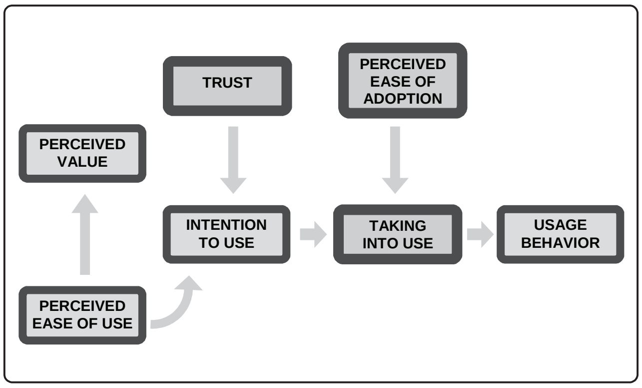
Table 1. The user evaluation activities that the technology acceptance model for mobile services was based on.

Service, application or device	Research methods	Users	Original results published in
WAP services	Laboratory evaluation with phone simulator	6	Kaasinen et al., 2000
WAP-converted Web services	Laboratory evaluation with phone simulator	4	
WAP services WAP-converted Web services	Field trial 2 months	40	Kaasinen et al., 2001
	Interviews with service providers	25	
WAP services WAP-converted Web services Web/WAP Message board for group communication	Field trial 2 months	40	
	Interviews with service providers	11	
Scenarios of personal navigation services	Group interviews	55	Kaasinen, 2003
Benefon GPS phone and services	Field evaluation	6	
Sonera Pointer location-aware WAP services	Laboratory evaluation	5	
Garmin GPS device	Field evaluation	5	
Magellan GPS device	Field evaluation	5	
Location-aware SMS services	Field evaluation	6	Kaasinen, 2005a
Weather and road conditions by SMS	Field trial, 1 month	10	
Location-aware integrated service directory	Field trial, 3 weeks	7	
Mobile topographic maps	Field evaluation	6	
Mobile 3D maps	Laboratory evaluation	6	
	Field evaluation	4	
Scenarios of context-aware consumer services	Interviews in anticipated contexts of use	28	

in their everyday lives and which features make the services acceptable in actual usage. As not all the field study findings could be fit to the original TAM model, it was necessary to update the model according to the repeated field study findings and themes identified in related research. The new model extends the original core model by Davis (1989) by identifying two new perceived product characteristics that affect the intention to use, that is trust and ease of adoption, and by redefining the theme of usefulness as value to the user.

The framework (Figure 5) suggests that perceived ease of use, perceived value, and trust affect the intention to use a mobile service. To get from an intention to use to real usage, the user has to take the service into use. This transition is affected by the perceived ease of adoption. Perceived value, perceived ease of use, trust and perceived ease of adoption need to be studied in order to assess user acceptance of mobile services.

Figure 5. Technology acceptance model for mobile services (Kaasinen, 2005b) as an extension and modification of TAM by Davis (1989)



The technology acceptance model for mobile services (Kaasinen, 2005b) constitutes a framework that helps designers of mobile services to identify key issues that should be focused on in the design to ensure user acceptance. Thus the motivation of the model is different than the motivation of the original TAM, which was built to explain user acceptance and underlying forces for existing technical solutions.

Perceived ease of use was included in the original TAM and it is also included in the TAMM model. Davis (1989) defined perceived ease of use as "*the degree to which a person believes that using a particular system would be free from effort*." At first, perceived ease of use is based on external factors such as the user's attitude towards technology in general, experiences of using similar services and information from other people. In actual use and sustained use, perceived ease of use is increasingly affected by the user's own experiences of using the system in different contexts of use.

In the case of mobile services that are used on small devices such as mobile phones or PDAs, the limitations of the device have a major influence on perceived ease of use. The limitations include the small screen, small and limited keyboard, the absence or limited functionality of pointing devices, limited amount of memory, limited battery power, and slow connections. As new devices and mobile networks are being introduced to the market, these limitations have somewhat diminished but still mobile networks are slower than fixed ones and the requirements for ease of carrying and holding the device do not allow very large screens or large keyboards. Designing mobile services for ease of use is to a large extent about coping with the limitations of the device. In addition, the design should adapt to the variety of client devices and available networks and other infrastructures.

The ease of use of mobile services has been studied quite a lot and different usability guidelines are available. It is a pleasure to note that many of the usability problems identified in early mobile Internet studies have already been corrected in current mobile devices, browsers and services. However, location-aware services pose even more challenges for ease of use. Location-aware services are not just mobile in the sense that they can be easily carried around but, typically, they are used while the user is moving. These kinds of usage situations require extreme ease of use. Personalization and context-awareness are expected to improve ease of use, but they may also introduce new usability problems, for example in the form of personalization dialogues.

Perceived value replaces perceived usefulness in the TAMM model because in our field trials with consumers it became evident that in the consumer market, perceived usefulness may not indicate adequate motivation to acquire the mobile service. As the focus group studies by

Järvenpää et al. (2003) point out, consumers may lack a compelling motivation to adopt new mobile services unless those services create new choices where mobility really matters and manage to affect people's lives positively. In a value-neutral setting each requirement is treated as equally important in the design (Boehm, 2003). This easily leads to featurism—the product becomes a collection of useful features but as a whole it may not provide enough value to the user. Value not only includes rational utility but also defines the key features of the product that are appreciated by the users and other stakeholders, that is the main reasons why the users are interested in the new product. As Roto (2006) points out, costs of using the service also affect the perceived value as user expectations tend to be higher for more expensive products. Values are made explicit by the identification of objectives, which are statements about what the user wants to achieve. Fundamental objectives are directly related to the user's current problem or situation at hand, whereas means objectives help to achieve the fundamental objectives (Nah et al., 2005).

Defining the targeted values and concentrating on them in design and evaluation helps to focus the design on the most essential issues. This is in line with the concept of value-centered software engineering proposed by Boehm (2003) and value-centered HCI proposed by Cockton (2004a, b). Focusing on perceived value in user acceptance studies supports the wider scope of value-centered design, where user value can be studied in parallel with business value and strategic value as proposed by Henderson (2005).

Trust is added as a new element of user acceptance in the TAMM model. The original TAM (Davis, 1989) was defined for information systems at work, and in those usage environments the end-users could rely on the information and services provided and the ways their personal data was used. When assessing user acceptance of e-commerce applications, Gefen et al. (2003) proposed to enhance TAM with trust in the service provider, as in their studies trust-related issues turned out to have a considerable effect on user acceptance. In our studies with mobile Internet, consumers were using mobile services that were provided to them via complex mobile service networks. In this environment trust in the service providers turned out to be an issue. As location-based services collect and use more and more information about the usage environment and the user, ethical issues arise. Especially ensuring the privacy of the user was a common concern of our test users. As the users get increasingly dependent on mobile services, reliability of the technology and conveying information about reliability to the user becomes more important.

In the technology acceptance model for Mobile Services, trust is defined according to Fogg and Tseng (1999). Trust is an indicator of a positive belief about the perceived reliability of, dependability of, and confidence in a person, object or process. User trust in mobile services includes perceived reliability of the technology and the service provider, reliance on the service in planned usage situations, and the user's confidence that he or she can keep the service under control and that the service will not misuse his or her personal data.

Perceived ease of adoption is related to taking the services into use. In the original TAM settings with information systems at work, this certainly was not an issue as users typically got their applications ready installed. In our field trials it turned out that a major obstacle in adopting commercial mobile services was the users' unawareness of available services, as well as problems anticipated in taking services into use (Kaasinen, 2005b). Furthermore, as usage needs were typically quite occasional, people often did not have enough motivation to find out about these issues. And finally, configuration and personalization seemed to require almost overwhelming efforts (Kaasinen, 2005b). Introducing the services to users would definitely require more attention in service design (Kaasinen et al., 2002).

As mobile services are typically used occasionally and some services may be available only locally in certain usage environments, ease of taking the services into use becomes even more important. The user should easily get information about available services and should be able to install and start to use the services easily. Finally, he or she should be able to get rid of unnecessary services.

Compared with the original TAM (Davis, 1989), the technology acceptance model for mobile services includes an additional phase between the intention to use and the actual usage behavior. Taking a service into use may constitute a major gap that may hinder the transfer from usage intention to actual usage (Kaasinen, 2005b). Perceived ease of adoption is added to the model at the stage when the user's attention shifts from intention to use to actually taking the service into use.

The characteristics of the user and his or her social environment affect how the user perceives the service. These issues are not included in the core TAMM model that aims to identify key characteristics of mobile services that generally affect user acceptance of mobile services. Further research is needed to fit previous TAM enhancements such as TAM2 (Venkatesh & Davis, 2000) and UTAUT (Venkatesh et al., 2003) to the model to identify external factors such as characteristics of the users and their social environment that affect the user acceptance factors in the model.

In the following section the technology acceptance model for mobile services is analyzed further and design implications are presented for each user acceptance factor, based on the synthesized results of the original case studies (Kaasinen, 2005b). The technology acceptance model for mobile services, together with the design implications, communicates previous user acceptance findings to the design of future mobile services.

DESIGN IMPLICATIONS

The technology acceptance model for mobile services defines four main user acceptance factors: perceived value, perceived ease of use, trust and perceived ease of adoption. How these factors should be taken into account in the design of individual mobile services depends on the service in question. However, there are many attributes of the acceptance factors that repeat from one service to another. These attributes form a set of design implications that can be used in the design of mobile services. The design implications can additionally be used in designing user acceptance evaluations to define the issues to be studied in the evaluation. In the following, design implications for each user acceptance factor are presented by combining results from the original studies (Kaasinen et al., 2000; Kaasinen et al., 2001; Kaasinen, 2003; Kaasinen, 2005a; Kaasinen 2005b) and results from related research.

Because of the quality of the case study material, the design principles cover best mobile information services targeted at consumers. For other kinds of services, the technology acceptance model for mobile services as well as the design implications can certainly be used as a starting point but they may need to be revised.

Perceived Value

Values define the key features of the services that are appreciated by the users and other stakeholders, that is the main reasons why the users are interested in the new services. Defining the targeted values helps in focusing the design on the most essential issues. Value is also related to the costs of using the service, and for commercial products the relationship of these two attributes should be studied, as proposed by Roto (2006). The following list gives some ideas about where in our studies the value was found.

Successful Service Content is Comprehensive, Topical, and Familiar

In the early days of mobile Internet, service providers often thought that small devices would require just a small amount of contents. Our studies showed that mobile users need access to all relevant information, as deep as they are ready to go, but the information has to be structured in such a way that the user can choose to get the information in small portions. Users appreciate comprehensive services in terms of geographic coverage, breadth (number of services included) and depth (enough information in each individual service).

Topical information is likely such that the mobile service is the best way to keep up to date with what is going on. In our field trials examples of successful topical content included weather forecasts, traffic information, news topics and event information. Topical travel information, for instance, does not just give timetables but informs about delays and traffic jams and recommends alternative routes.

The user may perceive the service as being familiar because it resembles other mobile services that he or she has been using or because it resembles the same service or brand in a different environment such as Web, TV or newspaper. For instance, in our mobile Internet studies teletext services converted from the Web were well accepted because of their familiarity. Familiarity was also related to the provider of the service, as test users pointed out that they preferred using, for instance, news services from a familiar and trusted service provider.

The Service Should Provide Personal and User-Generated Content

Personalization is not just about selecting services and contents within services but also about making the user's own personal items available, as illustrated in Figure 6 by the setup of the personal mobile Internet pages in our trials. Also with location-based services the users appreciated the possibilities to complement, for example map data with their own information such as important places, favorite routes, and self-written notes.

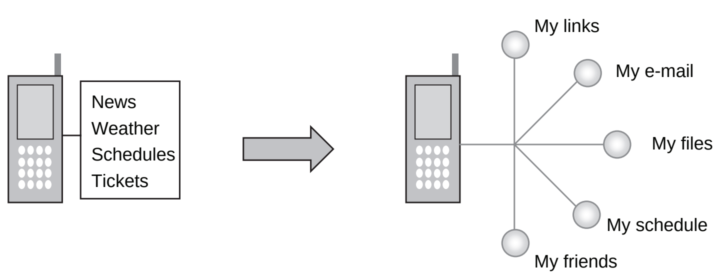
In the mobile Internet trials many users were keen to use services such as discussion groups where they could contribute as content providers. Letting the mobile users contribute to content creation could enhance many services. Such content may enrich the service, bring in additional users and encourage a sense of community among users. For instance, information generated by users at a particular location may be of interest to the next visitors. With the growing trend of social media services, the role of users' own content generation is expected to become increasingly important. Mobile users have key roles in many social media services as they can contribute by bringing in topical information from the field, such as mobile video of important occasions.

The Users Appreciate Seamless Service Entities Rather than Separate Services

In the mobile Internet trials it turned out that usage needs for many individual services were quite occasional, even if the users would have assessed the services as being very useful in those occasional situations. The value of mobile Internet to the user was based on the wide selection of services rather than any individual service.

The studies with location-based services pointed out the need for seamless service entities, whereby the user is supported throughout the whole usage situation, for example while looking for nearby services, getting information on the services, contacting the services, and getting

Figure 6. The shift from common to personal increases the appeal of the services



route guidance to find those services. The usage may even extend from one terminal device to another.

The Services Need to Provide Utility, Communication or Fun

In addition to personally selected content, interactive services also take mobile services to a more personal level, providing the users with new ways of communicating and participating. A mobile phone is basically a communication device and thus it is no wonder that services that enhanced or enriched communication were well accepted in our field trials.

Location-awareness can provide the users with services that are really intended for mobile use, not just secondary access points to Web services. Examples of such services include traffic information, weather forecasts, route guidance, travel information, event information and help services in emergency situations. Those services turned out to be popular as location-awareness made them both easier to use and more personal.

Perceived Ease of Use

Many ease of use attributes are already well known but as mobile services are getting increasingly complex and enhanced with new characteristics such as personalization and context-awareness, new usability challenges are raised. Key design principles that in our trials turned out to affect the perceived ease of use of mobile services are described in the following.

Clear Overview of the Service Entity

The most common usability problem with both mobile Internet services and location-based services was that when accessing the services, the users did not know what to expect from the service. The users would need a clear and intelligible overview of the whole range of available information, services and functions. The first impression may encourage and motivate the user or frighten him/her away. Enough design efforts and user evaluations should be invested in designing the main structure and the front page of the service. There are already efficient solutions available such as Minimap introduced by Roto (2006). Minimap gives the overview by showing a miniaturized version of the original Web page layout on the mobile device, and Roto's (2006) studies showed that this approach clearly improved the usability of Web browsing with mobile phones.

The information and functions that the user will most probably need should be the easiest to access. By proceeding further, the user should be able to access any information available within the service. Occasional usage typical of mobile services emphasizes the need for a clear overview of available services, including information on how the service should be used, where the content comes from, how often it is updated, and how comprehensive it is.

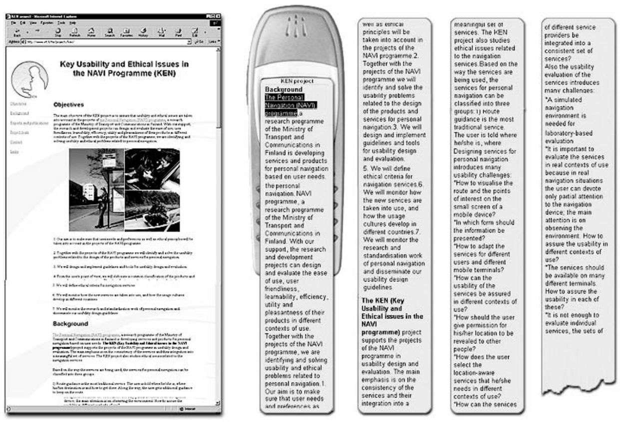
Fluent Navigation on a Small Screen

The mobile Internet trials showed that a single scrollable page (Figure 7) is good for browsing through information, whereas separate pages are better for navigation. The users need ways to browse quickly through less interesting information: for instance, an adaptive scroll speed and an illustrative scroll bar are useful.

The user needs clear feedback on which service and where in it he or she currently is. In our evaluations, this was facilitated by descriptive and consistent link/page header pairs for back, forward, exit, home and other safe heavens within the service.

The usability of the sites can be further improved by making the structure adaptive according to each user. A novice user may want to get instructions first, whereas more experienced users may want to go straight into the service. For frequent users, the structure could be adaptive so that the most recently or most often used items are easily available.

Figure 7. On a small screen, there is a lot to scroll, even when accessing a simple Web page



Smooth User Interaction with the Service

User interface restrictions of mobile devices and the implementation of the user interface elements may hinder smooth user interaction. In our trials text input was often a major effort for the users, especially if the usage took place while moving. Still the users needed and wanted to give input to the services. In the mobile Internet evaluations ready-made selection lists turned out to be useful when the user was getting acquainted with the service, whereas experienced users often preferred text input. Preferably, both alternatives should be available. Text input should be predicted and interpreted to suggest corrections to possible misspellings. Location-awareness as such could also be utilized in text input, as suggested by Ancona et al. (2001). For instance, nearby streets or the most popular search terms at a certain location could be suggested to the user. Even though user input may be tedious, it should not be avoided in the services. However, the user should not be obliged to input information that is not absolutely necessary.

Personally Relevant Services and Information without Expending Effort on Personalization Set-up

Our trials repeatedly showed that users were not willing to do much regarding personalization, although they would have appreciated the personalized solution. Personalization should be voluntary, and strongly supported in the beginning.

Users could be provided with ready-made service package alternatives, as we did in our mobile Internet trials, or they could be guided through personalization services. New service offerings could be sent automatically based on user profiles if the user accepts that. The user should be able to see and refine the personalization with his or her mobile device on the fly, even though the personalization could also be done with a desktop PC. New approaches such as group profiles, profiles shared by several services and learning profiles may ease personalization.

Easy Access to Situationally Relevant Information and Services

Mobile contexts vary a lot and may even change in the middle of a usage session. Our trials with location-based services showed that in services targeted at a limited area, such as travel guides, service catalogues and event guides, the context of use can be predicted quite well according to user location and time. This gives possibilities for different context-aware features in the services, easing their use and giving the users personalized access to the services.

Location-awareness can be utilized to provide the users with local services such as tourist guides, event information and shopping guides. Context-awareness can be complemented with personalization to adapt to user preferences that in different contexts may vary from one individual to another. This may, however, indicate lots of personalization efforts.

Facilitating Momentary Usage Sessions on the Move

On the move the users can devote only part of their attention to using the service while their main attention is on their main task of moving. In our trials with location-based services, on-the-move use was typically non-continuous. A user could, for instance, activate a route guidance service and start using it but occasionally he or she had to put the device aside and do something else. Later on he or she returned to the service. For these kinds of usage sessions task resumability should be supported both in the terminal device and in the services. Pousman et al. (2004) point out that resumability can be supported, for example by atomic interaction sessions, by appropriate time-outs on unfinished operations, and by a stateless interaction model. The users should be able to use the services both on and offline.

Design for Device and Network Variety

One of the main challenges in designing mobile services is the growing variety of mobile devices, networks and other infrastructures. The Design for All approach (EDeAN, 2007) with regard to mobile services requires taking into account all kinds of devices, not just the most advanced ones. In our development work on mobile Internet services we found that a good starting point is a simple service, suitable for any device. The usability and the attractiveness of the service can then be improved by utilizing the unique features of each device in separate implementations. Our experiences from mobile Internet trials show that in mobile environments there may be needs for adaptive search services that would not only look for particular content, but also take into account the current client device. The search results could be prioritized according to how suitable the content is for the device and network that the user is currently using.

Trust

In the TAMM model, user trust in mobile services is quite a wide concept that includes perceived reliability of both the technology and the information and functions provided, reliance on the service in planned usage situations, and the user's confidence that he or she can keep the service under control and that the service will not misuse his or her personal data. The design principles that in our evaluations turned out to affect user trust in mobile services are described in the following.

The User should be able to Rely on the Service in Intended Contexts of Use

In our user trials errors with mobile services were often difficult to cope with for the users as they did not know whether the problems were in the mobile device, in the network or in the services. Repeated malfunctions that the user could not understand or solve were a major source of bad usage experiences and often made the user stop using the service in question. To avoid these kinds of situations, the user should get easy-to-understand information to help him or her to understand and recover from the error situation. User errors should be prevented by all means, for example by trying to interpret, correct or complete user input. In the event of the user losing the connection to the service, it should be assured that no harm will be done.

With location-based services the users often would have liked to get feedback on the power still available and estimates of the sufficiency of batteries with different combinations of add-on devices and functions (Kaasinen, 2003). A user on the move may need to make decisions regarding which combination of functions he or she can afford to keep on in order to avoid exhausting the battery power totally.

Evaluations of personal navigation scenarios and prototype services revealed that users may get quite dependent on mobile services such as navigation services. That is why the users should be made aware of the possible risks of using the product and they should be provided with information about the reliability of the service so that they can assess whether they can rely on the service in the planned usage situations.

Measurement without Estimated Accuracy is of no Use

The accuracy of the location information was often questioned in our trials. In addition to location, future mobile services will be using and providing the user with increasing amounts of different measurement data (Kaasinen et al., 2006). Accuracy requirements for the data need to be considered in the design. The accuracy should be sufficient for the kinds of tasks for which the user will be using the service. The users should get feedback on the freshness of the data and its accuracy, especially if these vary according to the usage situation. Both actual reliability and perceived reliability need to be ensured in the design as these may be only loosely mapped, as found out by Kindberg et al. (2004).

Context-aware systems have several error possibilities: the system may offer the user wrong things either because it predicted the context wrongly or because it predicted the context correctly but predicted the user's needs in that context wrongly. Displaying uncertainty to the user may improve the acceptability of the services by making them more intelligible, as pointed out by Antifakos et al. (2004).

The Privacy of the User must be Protected Even if the User would not require it

User data should be protected even if—like in some of our trials—the users themselves would be trusting enough not to require it. The user should be provided with easy mechanisms for giving permission to use the data for a predefined purpose. Histories of user data should not be stored purposelessly and without user consent. When location data is conveyed to others, it is worth considering whether they will need the exact location coordinates or a more descriptive but less intrusive description. It should also be considered whether it is necessary to connect personal data to the user identity.

The legislation in most countries requires the user's permission before he or she can be located. Also social regulation can create rules and norms for different situations in which location-aware services are used (Ackerman et al., 2001). In practice, trade-offs between privacy protection and effortless use need to be resolved.

In future services, it can be expected that in addition to user location, a lot of other personal data may be collected. This may include health-related measurements, shopping behavior; services used and so on (Kaasinen et al., 2006). The same principles as with location are to great extent valid also with this data.

The User Needs to Feel and Really be in Control

The more complicated the mobile services and the service networks behind them get, the less possibilities the user has to understand what is happening in the service. The services need to be somewhat seamless to ensure effortless use. On the other hand, some issues need to be clearly differentiated so as to ensure that the user understands what is going on. Seamless services may hide details from users when aiming to provide ease of use. This may prevent the user from understanding what is happening "behind the scenes" (Höök, 2004).

Based on the findings of the trials with location-aware services, the main user requirement is that the user needs to feel and really be in control. For instance, the users more easily accepted context-aware behavior of the services if they could understand the reason for the behavior. To be able to be in control, the user needs to understand enough about the system's capabilities and rules of reasoning. The user needs to get feedback on what is going on and why, even if it is unnecessary to understand all the details. As automated functions may take control away from the user, the user should be able to control the degree of automation and intrusiveness. The user should be able to override the recommendations of the system, as suggested by Cheverst et al. (2000).

Similar to the findings by Cheverst et al. (2002), also in our trials the users tended to accept push services because of the effortless use. However, as the amount of push features grows, the attitude of the users may soon change. That is why the user should be able to fine-tune or cancel the push feature easily—ideally as he or she receives a push message.

Perceived Ease of Adoption

As mobile services will increasingly be available from different sources and in complex service networks, it becomes important to ensure that the users get reliable information about available services and the necessary guidance when taking the services into use. Based on user feedback in our trials, key design principles regarding ease of adoption of mobile services are described in the following.

Real Values of the Services Need to be Emphasized in Marketing

Users often have a poor understanding of mobile devices and services (Kolari et al., 2002). The users may have misconceptions about the services behind acronyms or different technologies. In our

Table 2. Trade description model for personal navigation products and services (Kaasinen et al., 2002)

Classification	*Trade description*
User	**Is this product/service suitable for me?** • Targeted specially at a certain user group • Targeted only at a certain group • Accessibility for disabled users
User goal	**What can I do with this product / service?** • Locate myself • Be located by other people • Locate other people • Track my property • Get route guidance • Find and use nearby services • Get help in emergency situations • Have fun
Environment	**Where can/cannot I use this product/service?**
Equipment	**What do I need to know about the technology?** • What kind of technology do I need to be able to use the service? • How compatible is this product/service with other products/services? • How accurate is the positioning? • To what extent can I rely on this product?
Service characteristics	**What specific features does this service include, what is the added value of this product compared with competing products or current ways to act?**

trials the users were often unaware of the features and services available on their personal phones.

As a part of our research work a Trade Description Model (Kaasinen et al., 2002) was set up to help consumers to compare different products and, on the other hand, to help service providers to describe their products in a consistent way (Table 2). Although the model was designed for personal navigation services, it is general enough to be adopted for the description of other mobile services as well. The trade description model can also be used as a checklist of issues to be covered when writing "Getting started" manuals.

Disposable Services for Occasional Needs

IBM has issued guidelines on how to design out-of-box experiences that are productive and satisfying for users (IBM, 2005). Ideally, the services should be installed on the user device at the point of sale, and the user should at the same time get personal usage guidance, but presumably this will be possible with only a few services.

In our trials with location-based services, the users often said that they wanted to have the services easily available when a spontaneous need for a certain service arose. Context-aware services pose additional challenges for taking new services into use. The services may be available only locally or in certain contexts. The user should be able to identify, understand and take into use these services easily while on the move. As the selection of available services grows, it will also become increasingly important to get rid of unnecessary services easily.

The Service has to Support Existing and Evolving Usage Cultures

Personal mobile devices should be designed to be both intuitive for first-time use and efficient in long-term use (Kiljander, 2004). This is true also with mobile services, which should be designed for gradual learning. New services shape the usage, but the usage should also shape the services (Norros et al., 2003). Existing and evolving usage cultures should be studied in parallel with the technology development to identify and support natural usage patterns. The design should fit in with the social, technical and environmental contexts of use, and it should support existing usage cultures. Ideally, the technology should provide the users with possibilities that they can utilize in their own way, rather than forcing certain usage models fixed in the design (Norros et al., 2003). Although the users will benefit from clear usage guidance, they should also be encouraged to discover and innovate their own ways to utilize new services.

FUTURE TRENDS

The current technology acceptance model for mobile services (TAMM) is based on studies with mobile Internet services and location-based information services targeted for consumer use. The identified user acceptance factors can be utilized in designing these kinds of services, but they can also be applied when designing other kinds of mobile services. In future visions, mobile devices are increasingly interacting with their environment and are transforming into tools with which the user can orient in and interact with the environment. As the user moves from one environment to another, the available services will change accordingly (Kaasinen et al., 2006). These kinds of services will require extreme ease of adoption, and, as the services will increasingly deal with personal data, the user's trust in the services will become an even more important user acceptance factor.

Further studies will be needed to study the mutual relations of the four user acceptance factors. As with the original TAM, the model can be enhanced by studying key forces underlying the judgments of perceived value, perceived ease of use, trust and perceived ease of adoption.

The technology acceptance model for mobile services was set up by analyzing and combining the results of several individual evaluation activities of different mobile services. When developing future mobile technologies and infrastructures, human-centered design can be expanded similarly. By synthesizing and generalizing the results of parallel research activities, key user acceptance factors and design implications for future service development can be identified.

The technology acceptance model for mobile services seems to have potential as a framework for ubiquitous computing applications as well. The model has already been successfully applied in connection with a project that aims to develop a mobile platform for ubiquitous computing applications that utilize wireless connections to sensors and tags (Kaasinen et al., 2006).

CONCLUSION

In this chapter, the technology acceptance model for mobile services has been introduced. According to the model, user acceptance of mobile services is built on three factors: perceived value of the service, perceived ease of use, and trust. A fourth user acceptance factor: perceived ease of adoption is required to get the users from intention-to-use to actual usage. Based on the technology acceptance model for mobile services, design implications for each user acceptance factor have been proposed.

Instead of implementing collections of useful features, the design of mobile services should be focused on key values provided to the user. The value of mobile services can be built on utility, communication or fun. Successful service content is comprehensive, topical and familiar, and it includes personal and user-generated content. The users appreciate seamless service entities rather than separate services. Ease of use requires a clear overview of the service entity, fluent navigation on a small display, and smooth user interaction with the service. The users should get personally and relevant services and information without needing to expend effort on personalization. The services should be designed to be adaptive to a wide variety of devices and networks. As the services increasingly support individual users in their daily tasks and increasingly deal with personal data, user trust in the services is becoming more and more important. The user should be able to assess whether he or she can rely on the service in the intended contexts of use. The user needs to feel and really be in control, and the privacy of the user must be protected.

Occasional usage and momentary usage sessions on the move are typical of mobile services. In addition, services are increasingly available only locally or in certain contexts of use. This indicates the need for disposable services: services that are easy to find, take into use, use and get rid of when no longer needed. The user needs realistic information about the actual values of the services, so that he or she can realize how to utilize the service in his or her everyday life and discover new usage possibilities.

The technology acceptance model for mobile services provides a tool to communicate key user acceptance factors and their implications to the design. The model can be used in all design and evaluation activities throughout the design process, but it is especially useful in identifying issues that should be examined in field studies.

REFERENCES

Ackerman, M., Darrel, T., & Weitzner, D. J. (2001). Privacy in context. *Human-Computer Interaction, 16*, 167–176.

Ajzen, I., & Fishbein, M. (1980). *Understanding attitudes and predicting social behaviour*. Prentice Hall.

Ancona, M., Locati, S., & Romagnoli, A. (2001). Context and location aware textual data input. *SAC 2001* (pp. 425-428). Las Vegas: ACM.

Antifakos, S., Schwaninger, A., & Schiele, B. (2004). Evaluating the effects of displaying uncertainty in context-aware applications. In Davies, N., Mynatt, E., & Siio, I. (Eds.), in *Proceedings of Ubicomp 2004: Ubiquitous Computing 6th International Conference* (pp. 54-69). Springer-Verlag.

Barnes, S. J., & Huff, S. L. (2003). Rising Sun: imode and the wireless internet. *Communications of the ACM, 46*(11), 79–84.

Billsus, D., Brunk, C. A., Evans, C., Gladish, B., & Pazzani, M. (2002). Adaptive interfaces for ubiquitous Web access. *Communications of the ACM, 45*(5), 34–38.

Boehm, B. (2003). Value-based software engineering. *Software Engineering Notes, 28*(2), 1–12.

Chen, L., Gillenson, M. L., & Sherell, D. (2004). Consumer acceptance of virtual stores: A theoretical model and critical success factors for virtual stores. *ACM SIGMIS Database archive, 35*(2), 8–31.

Cheverst, K., Davies, N., Mitchell, K., Friday, A., & Efstratiou, C. (2000). Developing a context-aware electronic tourist guide: some issues and experiences. *CHI 2000 Conference Proceedings* (pp. 17–24). ACM.

Cheverst, K., Mitchell, K., & Davies, N. (2002). Exploring context-aware information push. *Personal and Ubiquitous Computing, 6*, 276–281.

Cockton, G. (2004a). From quality in use to value in the world. In *Proceedings of CHI2004* (pp. 1287-1290). ACM.

Cockton, G. (2004b). Value-centred HCI. In *Proceedings of the Third Nordic Conference on Human-Computer Interaction* (pp. 149–160).

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quartely, 13*, 319–339.

Davis, F. D., & Venkatesh, V. (2004). Toward preprototype user acceptance testing of new information systems: implications for software project management. *IEEE Transactions on Engineering Management, 51*(1).

EDeAN. (2007). European design for all e-accessibility network. *Homepage*. Retrieved February 13, 2007, from www.e-accessibility.org

Fogg, B. J., & Tseng, H. (1999). The elements of computer credibility. In *Proceedings of CHI 99 Conference* (pp. 80–87).

Funk, J. L. (2004). *Mobile disruption. The technologies and applications driving the mobile Internet.* Wiley-Interscience.

Gefen, D. (2000). E-commerce: The role of familiarity and trust. *Omega: International Journal of Management Science, 28*, 725–737.

Gefen, D., & Devine, P. (2001). Customer loyalty to an online store: The meaning of online service quality. *Proceedings of the 22nd International Conference on Information Systems* (pp. 613–617).

Gefen, D., Karahanna, E., & Straub, D. W. (2003). Inexperience and experience with online stores: The importance of TAM and Trust. *IEEE Transactions on Engineering Management, 50*(3), 307–321.

Henderson, A. (2005). Design: The innovation pipeline: Design collaborations between design and development. *ACM Interactions, 12*(1), 24–29.

Höök, K. (2004). Active co-construction of meaningful experiences: but what is the designer's role? *Proceedings of the Third Nordic Conference on Human-Computer Interaction* (pp. 1-2). ACM Press.

IBM. (2005). Out-of-box experience. Retrieved January 10, 2005, from www-3.ibm.com/ibm/easy/eou_ext.nsf/publish/577

ISO 13407. (1999). *Human-centred design processes for interactive systems.* International standard. International Standardization Organization. Geneve.

Järvenpää, S. L., Lang, K. R., Takeda, Y., & Tuunanen V. K. (2003). Mobile commerce at crossroads. *Communications of the ACM, 46*(12), 41–44.

Kaasinen, E. (2003). User needs for location-aware mobile services. *Personal and Ubiquitous Computing, 6*, 70–79.

Kaasinen, E. (2005a). User acceptance of location-aware mobile guides based on seven field studies. *Behaviour & Information Technology,24*(1), 37–49.

Kaasinen, E. (2005b). *User acceptance of mobile services—Value, ease of use, trust and ease of* *adoption.* Doctoral dissertation. VTT Publications 566. Espoo: VTT Information Technology.

Kaasinen, E., Aaltonen, M., Kolari, J., Melakoski, S., & Laakko, T. (2000). Two approaches to bringing internet services to WAP devices. *Computer Networks, 33*, 231–246.

Kaasinen, E., Ermolov, V., Niemelä, M., Tuomisto, T., & Välkkynen, P. (2006). *Identifying user requirements for a mobile terminal centric ubiquitous computing architecture*. FUMCA 2006: System Support for Future Mobile Computing Applications. Workshop at Ubicomp 2006.

Kaasinen, E., Ikonen, V., Ahonen, A., Anttila, V., Kulju, M., Luoma, J., & Södergård, R. (2002). *Products and services for personal navigation—Classification from the user's point of view.* Publications of the NAVI programme. Retrieved January 4, 2005, from www.vtt.fi/virtual/navi

Kaasinen, E., Kasesniemi, E.-L, Kolari J., Suihkonen, R., & Laakko, T. (2001). Mobile-transparent access to web services—Acceptance of users and service providers. In *Proceedings of International Symposium on Human Factors in Telecommunication.* Bergen, Norway

Kiljander, H. (2004). *Evolution and usability of mobile phone interaction styles.* Doctoral thesis. Helsinki University of Technology. Publications in Telecommunications Software and Multimedia. TML-A8. Espoo: Otamedia.

Kindberg, T., Sellen, A., & Geelhoed, E. (2004). Security and trust in mobile interactions—A study of users' perceptions and reasoning. In Davies, N., Mynatt, E., & Siio, I. (Eds.), *Proceedings of Ubicomp 2004: Ubiquitous Computing 6th International Conference* (pp. 196-213). Springer-Verlag.

Kolari J., Laakko T., Kaasinen E., Aaltonen M., Hiltunen T., Kasesniemi, E.-L., Kulji, M., & Suihkonen, R. (2002). *Net in pocket? Personal mobile access to web services.* VTT Publications 464. Espoo: Technical Research Centre of Finland.

Mathieson, K., Peacock, E., & Chin, W. W. (2001). Extending the technology acceptance model: The influence of perceived user resources. *The DATA BASE for Advances in Information Systems, 32*(3), 86–112.

Nah, F. F.-H., Siau, K., & Sheng, H. (2005). The value of mobile applications: A utility company study. *Communications of the ACM, 48*(2), 85–90.

Norman, D. A. (1998). *The invisible computer.* Cambridge, MA: MIT Press.

Norros, L., Kaasinen, E., Plomp, J., & Rämä, P. (2003). *Human-technology interaction. Research and design. VTT Roadmap.* VTT Research Notes 2220. Espoo: Technical Research Centre of Finland.

Pousman, Z., Iachello, G., Fithian, R., Moghazy, J., & Stasko, J. (2004). Design iterations for a location-aware event planner. *Personal and Ubiquitous Computing, 8,* 117–125.

Rogers, E. M. (1995). *The diffusion of innovations.*(Fourth Ed.). New York: Free Press.

Roto, V. (2006). *Web browsing on mobile phones—Characteristics of user experience.* Doctoral dissertation. Espoo: Helsinki University of Technology.

Sarker, S., & Wells, J. D. (2003). Understanding mobile handheld device use and adoption. *Communications of the ACM, 46*(12), 35–40.

Venkatesh, V., & Davis, F. D. (2000). Theoretical extension of the technology acceptance model: Four longitudinal field studies. *Management Science, 46*(2), 186–204.

Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS Quarterly, 27*(3), 425–478.

KEY TERMS

Ease of Adoption (TAMM): Perceived ease of identifying, understanding and taking into use new products.

Innovation Diffusion: User adoption of different innovations in target populations

Location-Aware Service: A special case of location-based service: a mobile service that adapts according to the location.

Location-Based Service: A mobile service that utilizes location data.

Perceived Ease of Use (TAM and TAMM): The degree to which a person believes that using a particular system would be free from effort (Davis, 1989).

Perceived Usefulness (TAM): The degree to which a person believes that using a particular system would enhance his or her performance in a certain task (Modified from Davis, 1989).

Technology Acceptance: User's intention to use and continue using a certain information technology product (Davis, 1989).

Technology Acceptance Model (TAM): Technology acceptance models aim at studying how individual perceptions affect the intentions to use information technology as well as the actual usage. The Technology Acceptance Model was originally defined by Davis (1989), but it has subsequently been modified and augmented by other researchers.

Technology Acceptance Model for Mobile Services (TAMM): Extension of the original Technology Acceptance Model to take into account the specific characteristics of mobile services (Kaasinen, 2005b)

Trust (TAMM): An indicator of a positive belief about the perceived reliability of, dependability of, and confidence in a product (modified from Fogg & Tseng, 1999).

Value (TAMM): The key features of the product that are appreciated by the users and other stakeholders, i.e. the main reasons why the users are interested in the new product (Kaasinen, 2005b).

Chapter VIII
Transgenerational Designs in Mobile Technology

Martina Ziefle
RWTH Aachen University, Germany

Susanne Bay
RWTH Aachen University, Germany

ABSTRACT

Mobile devices have proliferated into most working and private areas and broad user groups have access to mobile technology. This has considerable impact on demands for usable designs. As users differ widely regarding age, upbringing, experience and abilities, it is a basic question whether there are user interface designs feasible that meet the demands of user diversity and trans-generational designs. The aim of the present research was to uncover effects of user diversity on menu navigation. Users of a wide age range were examined when interacting with mobile phones. In a detailed way, individual navigation routes were analyzed and effectiveness and efficiency of menu navigation was determined. In addition, effects of individual variables were considered. The results show that the usage of small-screen devices imposes considerable difficulties for all users, but in particular for children and middle-aged adults, who were very sensitive for cognitive demands imposed by current mobile phone designs.

INTRODUCTION

The distribution of mobile devices represents one of the fastest growing technological fields ever. Especially, small interface devices are omnipresent and can be characterized as important technical devices in today's societies. Mobile devices promise to be ubiquitously applicable and cover basic communication as well as office functionalities and allow Internet access. Moreover, the devices are used for route and traffic information, but provide also fun and entertainment applications.

The ubiquity and penetration of mobile devices raise new usability concerns. Many users show considerable problems with respect to the handling, learning, and understanding of these devices, which in turn reduce the ease of use and the perceived usefulness (e.g., Arning & Ziefle, 2006, 2007; Jakobs, 2005; Tuomainen & Haapanen; 2003; Ziefle & Bay, 2004; 2005). Yet, commonly agreed rules, which complexity of functions and which interface design is appropriate, have not been defined, and perhaps due to this fact, usability is not an issue that manufacturers are primarily investing in.

Several factors can be referred to that contribute to these difficulties. While formerly the usage of information technology was mainly restricted to technology-prone users, today, all user groups are addressed by technology. The diversity of the target groups, however, requires a basic understanding of the human factor and should be adequately addressed by device design. Users differ considerably with regard to their needs, motivation, competencies, and aptitudes, which is reflected in users' age, gender, and experience with technical devices.

In addition, more and more transactions include the utilization of technical devices and demand the acceptance and the competence of using technical devices. Thus, technical device usage is increasingly less optional, but represents more and more an indispensable qualification for many working settings. Furthermore, the nature and number of the devices' functionalities is elementarily changing. The traditional functionality of mobile phones, making calls, is only one among many other functions and the devices have an increasing complexity. Aggravating, numerous different device types within and across brands can be found on the market. While the applications and functions are increasingly merging across device types, though, devices differ considerably with respect to their basic structure and interface design. Within cross-platform-designs, it is thus difficult to understand, which operation modes and "device logic" is specific for a certain device and which is valid across devices (e.g., Ziefle, Arning & Bay, 2006). Finally, the miniaturization of the devices also contributes to cognitive difficulty when using technology. The tiny devices have small keys and miniature displays, thus the key handling and the visibility of the displayed information is considerably complicated. Furthermore, due to the restricted display, only few functions can be seen at a time. This increases memory load, as users have to remember function names and their menu location. Also, spatial orientation in the menu is problematical. Users do not experience how the menu is structured and how many functions are in the menu. As a consequence, users often lose their way in the menu.

BACKGROUND

The development of mobile technology and the device interface design still seems mainly to concentrate on what young and experienced users want (Maguire & Osman, 2003). However, children (mobileyouth.org, 2005) and older adults (Arning & Ziefle, 2007; Ziefle & Bay, 2006) are now also major user groups and, though, have not been considered adequately so far. This may be due to the fact that there is only little knowledge whether these groups have specific difficulties when using small-screen devices, and also regarding the factors, which might hamper or benefit a purposeful interaction with these devices (e.g., Arning & Ziefle, 2006, Tuomainen & Haapanen; 2003). Instead, a lot of preconceptions are prevailing. According to casual comments of many participants in our lab, we experienced that there is a "common knowledge" about aptitudes and abilities of age groups interacting with technical devices. Older adults are assumed to be the taillight regarding technical competence (and interestingly, they characterize themselves the same way), and quite low interest in technical developments is ascribed to them. As they have a different upbringing and were educated in times when technical devices were far less complex, they are thought to be considerably penalized. Conversely, children are supposed to easily master the interaction with technical devices. They are believed to understand the mode of operation of those devices much faster by virtue of their contact with interactive technology (e.g., computers, video games) from early on.

Additionally, children's fascination for explorative and inquisitive activities is well known, therefore they are assumed to be especially qualified for the interacting with technical devices.

Contrary to these statements, it was found that adults and children show a similar performance when using technical devices. Both were very sensitive to the demands imposed by the devices and showed considerable performance losses in sub-optimally designed interfaces (e.g., Bay & Ziefle, 2005; Ziefle & Bay, 2004; 2005; 2006; Ziefle, Bay & Schwade, 2006). But the exclusive focusing on users' age for technical performance is not sufficient. Rather, age must be characterized as the carrier of individual characteristics that are known to affect technical performance: cognitive abilities, attitudes, gender, or computer experience. Therefore, we need to understand the interrelation of these factors. If we want to learn if there are designs feasible, that are suited for all user groups or if we want to identify shortcomings, we also need to understand the specific impacts of individual variables, and their interaction with age and gender. The knowledge of the factors, which might underlie the aging and gender impact, though, is mostly limited to the examination of adults. Moreover, the interplay of different factors and performance has not been investigated satisfactorily so far.

Among the individual variables, which are known to play a role for adults' menu navigation performance, spatial ability is very prominent. Persons with high spatial abilities outperformed those with lower levels of spatial ability (e.g., Arning & Ziefle, 2007; Egan, 1988; Goodman et al., 2004; Kim & Hirtle, 1995; Vicente, Hayes & Williges, 1987; Westerman, 1997; Ziefle & Bay, 2006). Also, (verbal) memory is essential for the performance in technical menus. Users with high memory abilities had a better orientation in the menu, because they better memorized the functions and menu locations (e.g., Arning & Ziefle, 2007; Bay & Ziefle, 2003; Hasher & Zack, 1988; Ziefle & Bay, 2006). Moreover, the gender factor is crucial, especially in combination with computer self-efficacy. Female users often show lower self-efficacy and higher computer anxiety (e.g., Busch, 1995 Davies, 1994; Downing, Moore & Brown,2005). Rodger and Pendharkar (2004) referred performance differences between women and men to differences in computer experience levels, which are often lower in women. The interrelation of gender effects and computer experience is corroborated by studies showing that playful and active exploring of technical menus are forming an incidental knowledge of the system, which in turn contributes to computer experience (e.g., Bay & Ziefle, in press; Beckwith et al., 2006). Interestingly, playful interacting with computer systems is a behavior that is more often observed in male than female users (e.g., Van den Heuvel-Panheizen, 1999).

MAIN FOCUS OF THE CHAPTER

Comprising, the usability of small-screen devices, is an important but sophisticated demand, especially when taking user diversity into account. The interplay of user characteristics for the performance in technical menus is complex and needs a detailed examination. This is what the present paper wants to contribute to. The main focus was directed to a detailed analysis of users' navigation performance when using mobile phones. To understand the impact of user diversity, children, younger, and middle-aged adults were examined. Furthermore, gender effects were explored. Also, the influence of previous experience with technology and cognitive factors, as well as motivational factors, were taken into account. On the basis of a detailed analysis of individual interaction patterns, some implications for the design of mobile phones are discussed.

Method

Independent and Dependent Variables

Two independent variables were examined. The first independent variable was users' age, comparing the navigation performance of children (9-10 years), younger (20-30 years) and middle-aged adults (40-61 years). The second independent variable was gender, comparing female and male participants. Furthermore, participants' experience with technical devices, their spatial ability and short-term memory capacity were determined and treated as between subject variables, possibly affecting performance when using mobile phones.

As dependent variables, the effectiveness and efficiency of navigation were determined. In order to get a detailed insight in navigation behavior and to identify individual navigation patterns, six different measures were surveyed.

For the task *effectiveness*, the percentage of successfully solved tasks (within the time limit of five minutes per task) was measured. A maximum of eight tasks (four tasks solved twice) were to be completed.

Efficiency: (1) The *time* needed to process the tasks was surveyed. However, the time is a rather unspecific measure (as it does not tell us what users actually do in the menu). Therefore, more specific measures were determined in addition. (2) The number of *detour steps* (steps executed in the menu that were not necessary when solving the task on the shortest way possible (3) The number of *hierarchical returns* to higher levels in menu hierarchy, indicating that users in the belief of having taken the wrong path go back to a known menu position, consequently re-orientating themselves. (4) The number of *returns to the top.* This measure was assumed to reflect utter disorientation, as users had to re-orientate by returning to the top menu level, beginning from scratch.

Experimental Tasks

Four typical and frequently used mobile phone tasks were selected. In total, a minimum of 47 steps was necessary to solve the four tasks. Participants had to:

1. Call a number (11 steps)
2. Hide one's own number when calling someone (14 steps)
3. Send a text message (11 steps)
4. Make a call divert to the mailbox (11 steps)

In order to determine learnability effects, the tasks had to be solved twice consecutively. The order of tasks in the two trials was held constant over participants.

Apparatus and Materials

For the mobile phone, a well-known mass model, the Siemens S45, was chosen. In order to experimentally examine the quality of users' menu navigation performance, it was of methodological importance to analyze individual navigation routes in detail, and controlling for confounding factors at the same time. Therefore, the phone was simulated as software solution, run on a PC and displayed on a touch screen (Iiyama TXA3841, with a touch logic by ELO RS232C). Figure 2 shows a snapshot of the emulated phone. The display size corresponded to the original size, but the chassis of the phone and the keys were enlarged in order to enable easy operation of them with the finger on the touch screen. Moreover, a logging software tool was developed, which enabled us to log any user interaction with the system. By this, the number and type of keys used, the functions selected, and the individual navigation routes taken through the menu could be reconstructed in detail.

Figure 1. Snapshot of the emulated mobile phone (Siemens S45)

Participants

In total, 108 participants (58 females, 50 males) volunteered for the study. They were divided into three age groups: In the children group, 22 girls and 14 boys participated (M = 9.4; SD = 0.7). From the 36 young adults, 18 were females and 18 males (M = 24.1; SD = 2.8). Finally, 36 middle-aged adults (18 females, 18 males) took part (M = 47.1; SD = 7.6). The children were in their fourth school year. In the younger group, students of different academic fields volunteered. Participants of the middle-aged group were reached by an advertisement in a local newspaper and had a wide educational range. It was instructed that the study aimed at an evaluation of the usability of mobile phones. The motivation to join the study was high.

Assessing Users' Characteristics Interacting with Navigation Performance

As it was a major aim of the study to learn how the three age groups were interacting with the mobile phone and, moreover, which user characteristics might be crucial for navigation performance, the participants were surveyed regarding their spatial ability, verbal memory and the experience using technical devices. Here, the frequency and the reported ease of using these devices were assessed. Moreover, participants' interest in technology was determined.

Assessing Spatial Abilities and Verbal Memory

Assessing spatial and memory abilities in the children group, two subtests of the HAWIK-R were carried out. In the test on spatial ability ("Mosaic Test") the experimenter showed the child a picture (Figure 2) and the child's task was to reproduce the picture using cubes having different patterns on each of the sides. A maximum of 26 points could be attained in this test.

The test on short-term memory required the children to verbally repeat a row of numbers read aloud by the experimenter. The test consisted of seven rows of between three and nine numbers, which had to be reproduced directly after. The children were given two trials to correctly reproduce each row. A maximum score of 14 points could be reached.

For the adult group, spatial abilities were assessed with the paper-folding test (Ekstrom et al., 1976) in an online version (http://www.lap.umd.edu/vz2). Each of the 20 items includes successive drawings of two or three folds made in a square sheet of paper. The final drawing shows a hole punched in the folded paper. Participants had to mentally rotate the paper from the folded into the fully opened form and to indicate which of a number of possibilities shows the correct drawing. The 20 items had to be solved within 180 seconds. In Figure 3, an example item of the paper-folding test is given.

To assess memory ability in the adult groups, the verbal memory test adapted (Bay & Ziefle, 2003) from the learning and memory test (Bäumler, 1974) was used. Fifteen Turkish words (unknown to German participants) were presented in succession for three seconds each. Directly after the presentation, participants had to recognize the target items among three distractors, each being phonologically or visually similar. The maximum score to be reached was 15. An example from this test is given in Figure 4.

Figure 2. Mosaic test (Hawik-R): The upper row represents the single cube sides. The lower row represents one of the spatial tasks. The spatial demand for the children was to mentally deconstruct the figure into single cubes and to mentally rotate and arrange the cubes according to the figure.

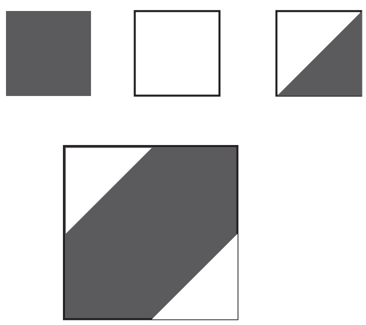
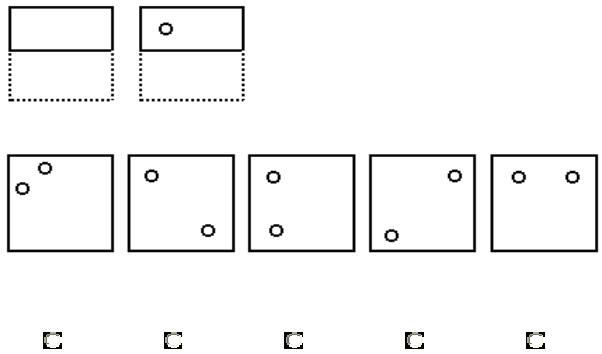


Figure 3. Item example of the paperfolding test (Ekstrom et al., 1976)



Assessing Previous Experience with Technical Devices

A detailed pre-screening of participants' experience with mobile phones and other technical devices as well as the reported ease of using them was carried out. Participants reported if and how often they use technical products (mobile phone, PC, video cassette recorder (VCR) and DVD), using a 5-point scale (1 = several times per day, 2 = once per day, 3 = once or twice a week, 4 = once or twice per month and 5 = less than once or twice a month). Furthermore, the estimated ease of using different technical devices had to be stated on a scale with four answering modes (1= the usage is easy, 2 = the usage is rather easy, 3 = the usage is rather difficult and 4 = the usage is difficult). Finally, participants indicated their interest in technology, using a 4-point scale (1 = low interest; 2 = rather low interest, 3 = rather high interest, 4 = high interest).

Results

The results were analyzed by multivariate analyses of variance assessing effects of age and gender on navigation performance in terms of effectiveness (number of tasks solved) and efficiency measures (time, detour steps, hierarchical returns, and returns to the top. Furthermore, the relationship between the users' age, gender and user characteristics was determined by correlation analyses. The level of significance was set at $p < 0.1$.

The result section is organized as follows. First the outcomes regarding user characteristics in the three age groups are described and how characteristics interrelate. Second, navigation performance is looked at, differentiating performance outcomes regarding age groups and also regarding gender effects. Then, learnability effects are focused upon, comparing the first contrasted to the second trial, determining if and to what extent performance improved for the three ages and for gender groups. A final analysis is concerned with the impact of user characteristics for performance outcomes.

Figure 4. Item example from the verbal memory test (Bay & Ziefle, 2003)

TATIL

☐ TAFIL ☐ TARAK ☐ TITAL ☐ TATIL

User Characteristics of Participants

In this section, the experience with technical devices, and the reported ease of using them are focused on. Also, the rated interest in technology is illustrated. Furthermore, the users' verbal memory and spatial abilities are described. It is of interest, if these variables are modulated by age or the gender of participants.

Previous Experience with Technical Devices

First, all participants reported to have a high experience using mobile phones. While younger and middle-aged adults possessed an own mobile phone, only 14 of the 36 kids did so. The children, who did not own a phone, though, reported to have frequent access to mobile phones (friends, siblings, or parents). Even though devices offer an increasing number of functionalities, astoundingly, only a very small fraction of these functionalities were used, across all age groups. However, there were age differences in the type of functionalities, which were commonly used. In the children group, the phones were mainly used for games (first choice) followed by calling and sending text messages. The young adult group indicated to use the phones mainly for text messaging and calling, but they also reported to play games and to use the phone as an alarm clock. Contrary, the middle-aged adult group reported to use the phones mainly for calling purposes, and emphasized that the majority of functions are "quite unnecessary." In Table 1, key results (means, standard deviations) are given for the experience measures and the interest in technology.

The frequency of mobile phone usage was significantly different for the three age groups (χ^2= 28.6; p = .000). The children and the middle-aged adult group used it 1-2 times a week (not differing from each other), while younger adults reported to use it once daily. With respect to PC usage, another significant age difference was found (χ^2 = 31.2; p = .000). The PC was used least in the child group (1-2 times a week), while both adult groups indicated to use it at least once a day. Finally, the frequency of using VCR/DVD is low, with the middle-aged adults using it about 1-2 times per month, while children and young adults use VCR/DVD about 1-2 times a week (χ^2 = 17.1; p = .000). When focusing on the ease of using these devices, all participants reported the usage as easy or at least rather easy. Age differences were

Table 1. Means (standard deviations) in user characteristics in all age groups

	Children	**Young adults**	**Middle-aged adults**
Gender	22 girls, 14 boys	18 women, 18 men	18 women, 18 men
Age	9.4 (0.7)	24.1 (2.8)	47.1 (7.6)
Frequency using a…	1= several times a day; 2 = Once daily; 3 = 1-2 times a week; 4 = 1-2 times a month; 5 = less than once a month		
Mobile phone	3 (1)	1.5 (0.8)	2.8 (1.8)
PC	3.2 (1.3)	1.5 (0.8)	1.9 (1.2)
VCR/DVD	2.9 (1.4)	3.2 (1)	4.2 (1.2)
Ease of using is…	1= easy; 2 = rather easy; 3 = rather difficult; 4 = difficult		
Mobile phone	1.4 (0.6)	1.2 (0.5)	2 (1.4)
PC	1.6 (0.7)	1.5 (0.6)	1.8 (1)
VCR/DVD	1.3 (0.5)	1.5 (0.7)	2.1 (1.4)
Interest in technology	1= low; 2 = rather low; 3 = rather high; 4 = high		
	3.4 (0.8)	2.8 (1.1)	2.4 (1.1)

found only with respect to VCR/DVD usage (χ^2= 8.8; p = .000). Finally, the interest in technology revealed another significant age difference (χ^2= 15.9; p = .000). The highest interest was present in the children group (M = 3.4). The lowest interest in technology was reported by the middle-aged adult group (M = 2.8), however, the young adults' interest was also comparably low (M = 2.4).

Across all age groups, female and male users did not differ regarding to the frequency of using mobile phones, VCR/DVD and PC, but introduced themselves as frequent users of common technical products. Also, no correlation of gender and the ease of using mobile phones and VCR/DVD devices were revealed. However, the interest in technology (r = -.26; p = 0.008) and the ease of using the computer (r= -.38; p = 0.000) showed significant correlations to gender: Female users reported the PC to be more difficult to use than the males and their interest in technology was lower compared to male users' interest in technology. Though interrelations were present in all age groups, interestingly, they were most pronounced for children.

Verbal Memory and Spatial Ability

First, outcomes in verbal memory in all age groups are addressed. On the left side of Figure 5, the scores of the children are illustrated. From the 14 points, the children reached, on average, "only" 5.4 points (SD = 1.5), and none of the children was able to reach the maximum score. Young adults (Figure 5, center) reached, on average, 12.4 points (out of 15). The middle-aged adult group (Figure 5, right) showed also a solid memory performance (M = 10.3; SD = 2.6), even though their memory score differed significantly from the younger adults' score (F (1,71) = 15.3; p=0.000).

In Figure 6, the outcomes in spatial abilities are pictured. The children (Figure 6, left) differed considerably with respect to spatial abilities. The inter-individual variance among children was high (range 4-26 points; M = 15.6; SD = 5.5), showing big developmental differences among 9-10 years old kids. For the younger group (Figure 6, center), the spectrum of correct answers ranged between 8 and 19 points (out of 20), reaching a mean performance of 13.2 (SD = 3.1). Finally, the middle-aged group reached an average score of 12.9 out of 20 points (SD= 3.7). The range of answers (4 points minimum and 20 points maximum) also represents a high variance, showing that spatial abilities do not follow a systematic decrease with increasing age. Statistical testing revealed no significant differences between spatial abilities of younger and middle-aged adults. Also, no gender differences were present neither with respect to verbal memory, nor spatial ability. However, for the children, there was a significant correlation of gender and the level of spatial ability (r = 0.6; p=0.03), with boys having higher spatial abilities (M = 18/26 points) than girls (M = 14/26 points).

Figure 5. Outcomes in verbal memory (left: children; center: young adults; right: middle-aged adults)

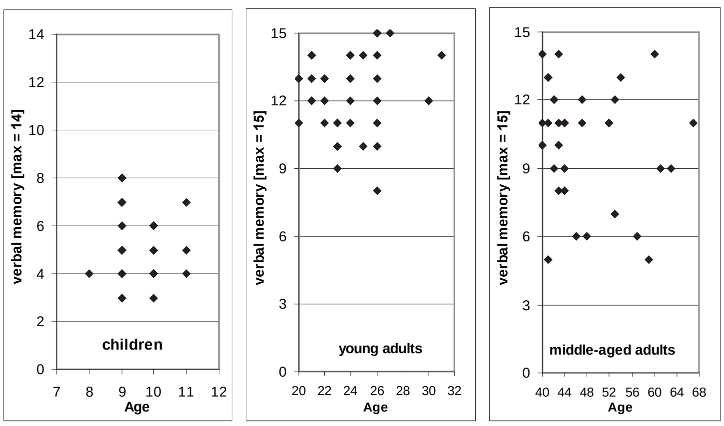
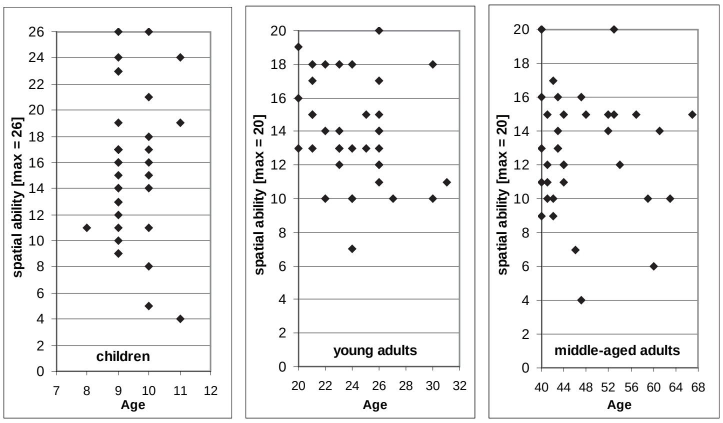


Figure 6. Outcomes in spatial abilities (left: children; center: young adults; right: middle-aged adults)



Menu Navigation Performance

When comprising the navigation performance in the eight tasks, a significant omnibus effect of age was found (F (2,102) = 18,1; p=0.00). Moreover, a significant omnibus effect of gender was revealed (F (1,102) = 3.1; p=0.01) as well as a significant interaction effect of age and gender (F (2,102) = 2.9; p=0.002). The age effect (single F-test; F (2,102) = 59.9; p<0.000) was based on significant differences between all age groups.

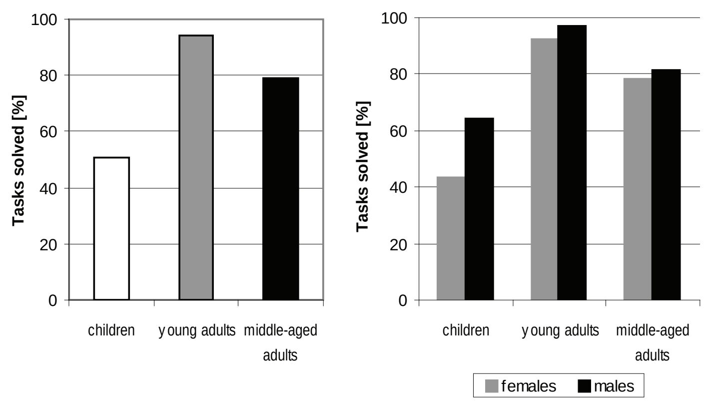
In Figure 7, the outcomes in task effectiveness are illustrated (left side: effectiveness for all age groups; right side; effectiveness for age and gender groups).

As can be seen there, the children showed the lowest effectiveness, reaching a mean task effectiveness of 51 percent. The best performance was present in the younger adult group, yielding a task effectiveness of 94 percent, while the middle-aged adults' effectiveness ranked in between (79 percent). Even though adult participants showed a considerably better performance than the children, it is quite astounding that not even the students were able to solve the eight tasks completely successfully. Also the effects of gender (F (1,102) = 11.3; p=0.001) as well as the interaction of age and gender (F (2,102) = 3.1; p=0.05) become obvious. The significant gender effect was originated by the girls' lower task success compared to the boys (girls: 43 percent; boys: 64 percent), while gender differences in the adults groups were not found to yield significant effects.

Furthermore, task efficiency is considered. Effects of age were significant for each of the single measures (time: F (2,102) = 50.5; p=0.000; detour steps: F (2,102) = 41; p<0.000); hierarchical returns: (F (2,102) = 22.6; p<0.000; returns to the top: (F (2,102) = 12.3; p=0.000). Independently of the measure, the children showed the lowest task efficiency, followed by the middle-aged adult group. The best task efficiency was present for young adults.

In Figure 8, the key results in task efficiency are illustrated for age groups and gender. As can be seen there, the children needed 20 minutes and 36 seconds and made 672.7 detour steps, when processing the phone tasks. The detouring of the children is considerable when taking into account that, overall, a minimum of 94 steps were needed to solve the tasks (47 steps per trial). Furthermore, the children made about 81.1 returns in menu hierarchy, and left the menu 8.1 times, to begin from scratch. Compared to the children, both adult groups were much more efficient even when the young adults significantly outperformed the middle-aged adult group. For the student group, it took 6 minutes 39 seconds to complete

Figure 7. Task effectiveness (%) in the three age groups (left) and both gender groups (right)



the tasks; compared to 14 minutes 19 seconds in the middle-aged group- this equals a benefit of 57 percent. Also, the students carried out only—but still—213.7 detour steps, returned 15.5 times to higher levels in menu hierarchy and began 0.9 times from the very beginning. In contrast, the middle-aged group made 414 detour steps, and returned, on average, 52.5 times to higher levels in menu hierarchy. Also, they made 5.4 complete returns to the top menu level.

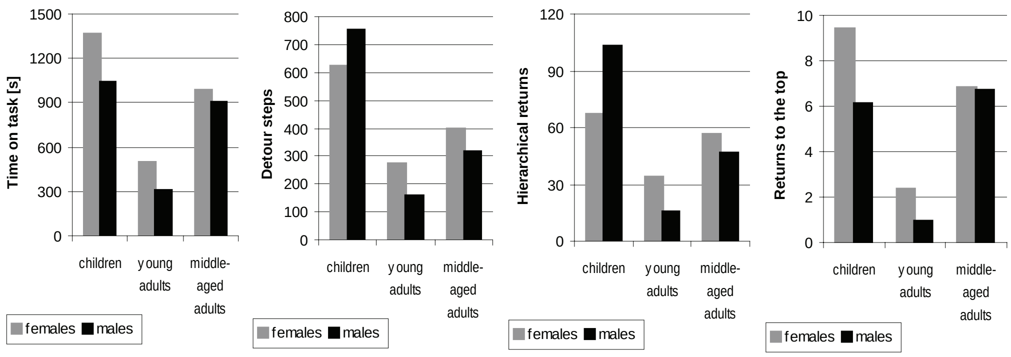
Looking on gender effects, male participants showed a generally more effective, and also a more efficient navigation style. However, gender effects were not symmetrical across ages and measures, respectively. The most pronounced gender differences were present in the children group, while gender effects decrease with increasing age. A quite interesting effect was revealed for the navigation strategy, which was different for girls and boys. The boys were more successful solving the tasks compared to the girls. Also, they needed less time and showed overall a smaller amount of returns to the top. So far, their navigation style is similar to adult male participants when compared to female participants. Yet, the boys' higher effectiveness was reached by a higher amount of exploration behavior in the menu. This can be taken from the high number of detour steps and hierarchical returns. Actually, their detouring was larger than that of all participants. Apparently, the boys capitalize the additional detouring on an overall better navigation performance, in contrast to the girls but also in contrast to both adult groups. Interestingly, the boys returned to a lesser amount to the top menu level compared to the girls, which shows from another side that the boys' detouring is more probably reflecting an active menu exploration rather than disorientation—otherwise they should have started from scratch more often.

Learnability Effects: Comparison of Navigation Performance in the First vs. Second Trial

Learnability effects, especially their interaction with age and gender effects might give additional insights in the difficulties users experience when interacting with small screen devices.

Again, first the task effectiveness is looked at. A significant learnability effect ($F(1,102) = 17$; $p=0.000$) was revealed, showing that in the second trial more tasks were solved successfully compared to the first trial (children: first trial: 47 percent, second trial 55 percent; young adults: first trial: 93 percent, second trial 95 percent; middle-aged adults: first trial: 75 percent, second trial 84 percent). No interaction effects of learnability and gender and learnability and age were revealed. Thus the higher task success in the second run was equally large for all participants.

Figure 8. Task efficiency (time on task, number of detour steps, hierarchical returns and returns to the top)



Next, efficiency is taken into account. In Figure 9 efficiency measures (time on task, detour steps, hierarchical returns and returns to the top) are depicted, differentiating the three age groups. Independently of the measure, there is a clear age pattern. Young adults showed the best performance and children the lowest performance.

According to the learnability effects regarding effectiveness, one would expect hat efficiency should also be higher in the second trial, therefore participants should spent less time on tasks, accompanied by fewer detour steps. In addition, also the number of returns in menu hierarchy and returns to the top should considerably decrease in the second run. Even though significant learnability effects for efficiency measures were present (time: $F(1,102) = 9.8$; $p=0.002$; detour steps: $F(1,102)=6,4$; $p = 0.01$), it becomes obvious that within navigation efficiency learnability effects are not equally high for all groups, but interact with the age of participants (time: $F(1,102) = 22.3$;$p=0.000$) as well as with gender (time: $F(1,102) = 7.8$; $p=0.006$; hierarchical returns: $F(1,102) = 5.5$; $p=0.002$; returns to the top: $F(1,102) = 2,5$; $p<0,1$). Moreover, there were also three-fold interactions between learnability, age and gender (time: $F(1,102) = 2, 3$; $p<0.1$; hierarchical returns: $F(1,102) = 2, 5$; $p<0.1$; returns to the top: $F(1,102) = 2, 4$; $p<0.1$). In order to disentangle the complex interrelation, first the nature of the interacting effect between learnability and age is addressed (Figure 9).

From Figure 9, it can be seen that actually only the young adult group profit from executing the tasks a second time. They were faster, executed less detour steps, and carried out fewer returns in menu hierarchy and also fewer returns to the top in the second trial compared to the first. The children and the older group, however, showed a different pattern. As the young adults, they were also faster in the second trial, however, in contrast to students, their detouring behavior did not improve in the second trial, as taken from the number of detours steps, the hierarchical returns and the returns to the top level. In short, one could characterize children's and older adults' navigation style as less cautious in the second compared to the first run (as they were faster), but still inefficient.

However, it is a basic question, whether learnability effects are similar across female and male users. This is analyzed in Figure 10. Here, tasks' efficiency is pictured for all participants, males (gray lines) and females (black lines) as well as for all age groups. The upper row of Figure 10 represents task efficiency of the children, the middle row navigation efficiency of the older group and the lower row shows efficiency measures of the young adults.

Again, we see the better overall performance of male compared to female users as well as the clear performance superiority of younger adults. Furthermore, it becomes obvious that the children group is considerably different compared to

Figure 9. Task efficiency (time on task, number of detour steps, hierarchical returns and returns to the top) in the first compared to the second trial for all age groups

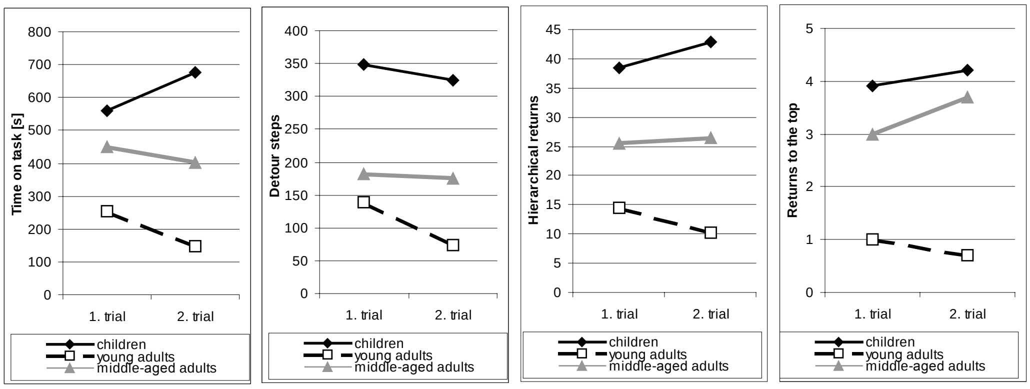
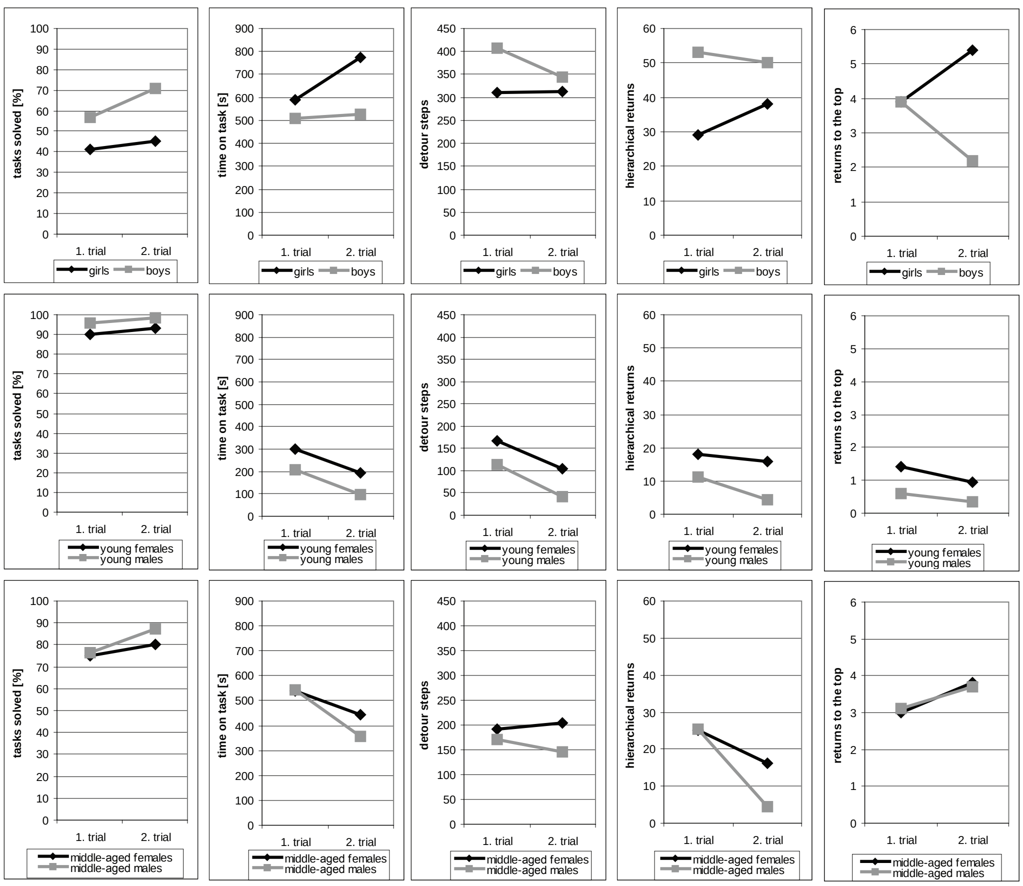


Figure 10. Task efficiency (time on task, number of detour steps, hierarchical returns and returns to the top) in the first compared to the second trial for all age groups (upper: children, middle: younger adults and lower row: middle-aged adults) as well as for female (black lines) and male (gray lines) participants



both adult groups: First, children had overall the lowest efficiency. Second, children did not profit from learnability, at least not the whole children group. Third, gender effects are most pronounced at this age, with boys having not only an overall better performance, but also revealing a different navigation style than the girls. The boys showed learnability effects, and profit from executing the tasks a second time, while, quite contrary, the girls did not show learnability effects and even had a lower performance in the second compared to the first run. Fourth, again, the boys' different navigation strategy becomes evident, which is characterized by a higher tasks success, a faster navigation with lower returns to the top. This effective strategy is reached by a more active and explorative menu navigation pattern—the boys carried out even more detour steps and hierarchical returns in the second trial than the girls in the first. With respect to the returns to the top, the measure for menu disorientation, the boys left the menu only twice in the second trial, compared to four times in the first.

User Characteristics and their Effects on Menu Navigation Performance

So far, we found performance differences between age and gender respecting effectiveness and efficiency when using a mobile phone. In this section it is analyzed, which of the user characteristics may account for these differences. Is users' experience with technical devices a substantial source of performance or is their interest in technology the source of the differences? Which role are the differences in spatial ability and verbal memory playing for menu navigation performance? First, the interrelations of the experience with technical device and the interest in technology are focused (Table 2).

Neither the reported interest in technology was interrelated with performance outcomes, nor the reported ease of using the devices showed significant correlations. Thus, motivational factors as the individual interest in technology were not decisive for performance. Also the perceived ease of using the devices did not reflect the actual navigation performance. However, participants' previous experience in terms of the self-reported

*Table 2. Correlations between user characteristics and navigation performance (***p = 0.000; ** p =0.05;* p = 0.1)*

N = 108	Tasks solved	Time on task	Detour steps	Hierarchical returns	Returns to the top
Frequency using a…					
Mobile phone	**r = -0.46*** **	**r = 0.5*** **	**r = 0.44*** **	**r = 0.45*** **	**r = 0.45*** **
PC	**r = -0.48*** **	**r = 0.41*** **	**r = 0.42*** **	**r = 0.36*** **	**r = 0.26** **
DVD/VCR	**n.s**	**n.s**	**n.s**	**n.s**	**n.s**
Ease of using a…					
Mobile phone	**n.s.**	**r = 0.27** **	**n.s.**	**n.s.**	**n.s.**
PC	**n.s.**	**r = 0.23** **	**n.s**	**n.s**	**r = 0.26** **
DVD/VCR	**n.s.**	**n.s.**	**n.s**	**n.s**	**n.s**

frequency of using technical devices was strongly interrelated with performance outcomes, showing that frequent usage and activities with the devices lead to a elaborated knowledge that is basically benefiting performance for all age groups (even though it should be considered that the experience level was not sufficient to reach an "optimal" performance (100 percent task success) in neither of the groups).

With respect to the impact of cognitive factors (spatial ability and verbal memory), correlation analyses were run for each age group, separately. The outcomes are summarized in Table 3.

For the children group, neither spatial abilities nor verbal memory had a significant impact for children's navigation performance. Thus, even those kids, who had—relatively to other children of this group—possessed a high verbal memory capacity and spatial ability were not specifically advantaged in menu navigation over those kids, who have only lower cognitive abilities. In other words: navigating through mobile phone menus was a high cognitive demand for all children. In contrast, both adult groups showed interrelations with spatial abilities, even though the relation was much more pronounced in the young adult group compared to the older adult group. Users with a high spatial score solved more tasks, were faster, and also made fewer detour steps and hierarchical returns. Also they did not return to the start level as often than adults with lower spatial abilities.

The impact of verbal memory capacity for navigation performance was comparably low. For the young adults, none of the performance measures showed significant interrelations with verbal memory. In the middle-aged adult group, however, effects of variability in memory capacity were revealed: users with high memory abilities solved significantly more tasks and were faster in comparison to those with a lower memory capacity.

Implications for Design

Even though the central focus of this work was directed to a detailed analysis of user behavior interacting with mobile phones and, also, to the impact of user diversity for performance, the present findings may also give some insights for design concerns to be considered for mobile devices. Also some training and tutoring issues may be derived.

The results reported here uncover both, similarities as well as differences in the navigation behavior of kids, young and older adults. Thus, we learn that there are design implications, which favor "a design for all approach" as well as differential aspects, which should be pursued to support specific user groups.

Across age groups, considerable difficulties were revealed in completing these common and easy phone tasks on a standard mobile phone. Not

*Table 3. Correlations between user characteristics and navigation performance (***p = 0.000; ** p =0.05;* p = 0.1)*

	Tasks solved	Time on task	Detour steps	Hierarchical returns	Returns to the top
Spatial ability					
Children	**r = 0.29***	n.s	n.s	n.s	n.s
Young adults	**r = 0.29***	**r = -0.36****	**r = -0.29***	**r = -0.48*****	**r = -0.29***
Middle-aged adults	**r = 0.37****	**r = -0.40****	**n.s**	**n.s**	**r = -0.45*****
Verbal Memory					
Children	n.s	n.s	n.s	n.s	n.s
Young adults	n.s	n.s	n.s	n.s	n.s
Middle-aged adults	**r = 0.40****	**r = -0.36****	n.s	n.s	n.s

even the students, bright and technology prone were able to solve the four tasks completely successfully. Nevertheless students showed the best performance compared to kids and older adults. All participants carried out a lot of detouring in the menu, as taken form the high number of detour steps, the returns in menu hierarchy and the returns to the top menu level, beginning from scratch. Thus, it must be concluded that current small screen devices are—cognitively—challenging to use. This is valid for kids and older adults, which—due to their developmental status—can be categorized as "weaker" users, but it is also valid for the "best case" student user group. According to statements of participants after the experiment, the difficulties they experienced in the menu had mainly three sources. The first difficulty referred to the complexity of the menu. Even though the task complexity was relatively low, with three menu levels at the most, participants (especially kids and older adults) had considerable problems to orientate, often not knowing were they were in the menu and where they had to go next. A second point refers to keys' complexity. Ambiguous functionality and design of keys lead to difficulties and provoked many unnecessary key actions. This was especially the case for navigation keys, which had a high complexity (keys, with several functions on different menu levels). Third, the naming of menus, sub-menus, and functions is also of crucial importance for good usability of a mobile phone.

However, there were also findings that hint at specificities of user groups, which should be considered. The first refers to different coping styles of children and older adults, when confronted with suboptimal and not very intuitive interfaces. Older users were nearly annoyed and insistently emphasized that they want to have devices that meet their demands of low complexity and all necessary functions within easy reach. Otherwise, they are not willing to use these devices. For the children, mobile phones still represent a high status and attractive gadget. Nevertheless, the children reacted highly sensitive to their failure and tended to attribute the failure to their own incompetence, but they also criticized that the mobile phone was "pretty hard" to use. This is of specific pedagogic impact. The success and the ease with which devices can be used contribute considerably to users' self-efficacy, the perceived competency and usability of technical devices (e.g., Arning & Ziefle, 2007; Ziefle, Bay & Schwade, 2006). Sub-optimally designed interfaces might lead to a lower frequency of using technical devices, in order to avoid negative feelings. However, frequent interaction with technical devices is an essential precondition for the formation of technical expertise, which in turn benefits navigation performance. Gender effects were also identified. Female users reported a significant lower interest in technology, and also, rated the ease of using technical devices as lower. While gender differences in performance did not reach significance level in both adult groups, for the children, they were most pronounced. The boys reached overall a higher performance, which was supported by a specific and successful navigation strategy. The boys were highly explorative and active in the phones' menu, pursuing a trial and error style. In contrast, the girls showed a much lower activity when interacting with the phone. Playful experimentation is assumed to yield educational benefits because the users may incidentally gain knowledge of the system by exploring its structure.

As practical implications from the current research, the following recommendations can be given:

Interface-design:

- Keep menu structures as flat as possible and avoid high complexity, this helps to reduce menu disorientation.
- Keys complexity should be held as low as possible. Avoid the allocation of many functions to single keys (multimode). Whenever keys with more than one function have to be used, those functions should be grouped to one key, which have a (semantically) similar meaning ("correct," and "step back"). Avoid mode keys with semantically dissimilar functions ("step back," "hang up"). Also, the spatial position of the keys is important. Very frequent functions should be allocated

to centrally located keys. Do not change position of frequently used keys.

- Respecting the naming of menu functions, basically, very similar terms ("phone setting" vs. "call setting") should be avoided. Also, do not use abstract terms ("incognito"), unfamiliar abbreviations ("GBG," "GSM") or technical terms ("D2 services," "SIM-activity"). For many users, they are not easy to understand and therefore, are not easily to be learned. Furthermore, very generic terms ("options," "settings") are highly misleading terms. Even though these terms are not difficult to understand, it is nevertheless not easy to deduce which functions are summarized under these category labels. Apparently, it is difficult to know which functions are and which are not present within these categories, as the general terms basically allow—semantically spoken—a targeted function to be housed, however, in many cases they actually do not. Categorizing the functions in intuitively understandable menus and sub-menus is also of great importance for a good design.

Training and pedagogic issues:

- It is essential to encourage active and playful interaction with the device quite early in the learning process, especially for children and middle-aged users, in order to enable the development of perceptions of achievement and competence. By playful interacting with technical devices, users get casually to know which features and functions are available, where these functions are located within the menu, and also how to activate the functions. This may especially benefit users, who cannot rely on a high spatial ability, a high verbal memory capacity or a high technical self-competence, to develop a solid expertise respecting using small screen devices. Thus, it is of high importance to motivate and encourage users to actively interact with mobile devices.

CONCLUSION

This study aimed at a critical actual inventory of user characteristics and the competence using a mobile phone. On the user side, effects of age and gender were analyzed, as well as technical experience and perceived ease of device usage. Moreover, interest in technology was taken into account. Further, spatial ability and verbal memory capacity were psychometrically determined and related to menu navigation performance. On the performance side, an elaborate analysis of navigation patterns was undertaken. Beyond the task success and the processing time, the individual extent of detouring behavior was analyzed. In this context we determined the number of detour steps, but also, how often participants returned to higher levels in menu hierarchy. As a measure for utter disorientation, we analyzed how often participants, after they delved into distraction, returned to the first menu level, beginning from scratch. Methodologically, this detailed analyzing procedure can be strongly recommended, as it mirrors exactly what users actually do and enables the determination of the relation between performance and user judgments, which are often biased. Considering that the majority of manufacturers evaluate mobile phones primarily operating with user ratings for evaluation purposes, the validity of user ratings is questionable. Of course, preference ratings can be obtained much more easily, but they possibly do not reflect the actual difficulties of users in the system. If a device is supposed to be accepted in the long run and also acknowledged by a diverse user group, the impact of a detailed analysis of navigation patterns seem essential.

Our participants had a solid experience with different state-of-the-art technical devices. It could be shown that the experience with technical devices considerably advantaged menu navigation performance. Interestingly, and perhaps contrary to expectations, the children had, relatively, the smallest computer experience, but nevertheless a high interest in technology—higher than the interest reported by both adult groups. As shown, the common prejudice of children to easily master the handling of technical devices—due to their early contact with technology—is not true, at least not in the most stringent form. From a pedagogic point of view, it is important to motivate especially kids to frequently use and handle technical devices, in order to support the formation of technical experience.

With respect to cognitive abilities, which underlie the performance in technical systems, spatial abilities turned out to be important for the menu performance. This shows that the navigating in small screen devices imposes considerable demands on users' ability to "spatially" orientate within the menu. Spatial abilities are assumed to provide a specific advantage. Persons with high spatial abilities are able to construct a mental representation of the systems' structure during navigation (Sein, Olfman, Bostrom et al., 1993), and therefore have a better orientation in the system. However, the impact of spatial abilities for performance turned out to be age-related. While younger and middle-aged adults were able to profit from high spatial abilities, this was not the case for the 9-10 year old kids in our experiment. This confirms earlier findings (e.g., Bay & Ziefle, in press; Shemakin, 1962), according to which the ability to cognitively process spatial hints and to mentally represent structural knowledge is fully developed not until children are 12-13 years. It would be insightful to examine if there are specific trainings or software tutors feasible, which can support younger children and help them to achieve a good performance.

A final remark is concerned with some limitations regarding the methodology used. Our results are based on laboratory experiments and on the interaction with a simulated mobile phone. This was accomplished in order to provide experimental control and to rule out confounding effects. However, we acknowledge that the results presented here might represent a solid underestimation of the real situation. In our experiment, the cognitive workload to use mobile devices was much lower than they usually are in the interaction with mobile devices in real environments. In a mobile context users have to manage different and complex demands, simultaneously, and in the laboratory setting a quiet setting was present and users were able to concentrate on the tasks. Also key handling and visibility problems may occur in real contexts, which were controlled for in the experiment. Another limitation refers to the selection of the middle-aged adult group, which is definitively not representative for the whole group of adult users, especially older users (65+). Thus, overall, we have to concede that the performance levels reached in our setting might be higher compared to more realistic settings.

FUTURE TRENDS

Due to the fast cycles of technical innovations and the development of novel and still more complex technical devices, usability demands will still increase. This is of vital interest facing the demographic change and the increasing prominence of mobile devices. Therefore, research activities should address user diversity more strongly than hitherto. Many topics in this context should be pursued in greater detail. One is to examine the nature and benefit of the exploratory behavior of users when interacting with technical devices. It is a central question if there are specific interaction strategies, which should be encouraged or supported by trainings, and if there are differential aspects, which should be applied in specific user groups. Another interesting research question is the question whether the findings reported here are limited to devices with an exclusively hierarchical menu structure (as the mobile phone) or if the navigation patterns found here can be transferred to devices with a network data structure, which might provoke a completely different interaction pattern.

REFERENCES

Arning, K., & Ziefle, M. (2006). What older users expect from mobile devices: An empirical survey. In *Proceedings of the IEA 2006*. Amsterdam: Elsevier.

Arning, K., & Ziefle, M. (2007). Barriers of information access in small screen device applications: The relevance of user characteristics for a transgenerational design. In C. Stephanidis and M. Pieper (Eds.), *User interfaces for all: Universal access in ambient intelligence environments* (pp. 117-136). Berlin, Germany: Springer.

Arning, K., & Ziefle, M. (2007). Understanding Age differences in PDA acceptance and performance. *Computers in Human Behavior, 23*(6), 2904-2927.

Bäumler, G. (1974). *Learning and memory test*. Göttingen, Germany: Hogrefe.

Bay, S., & Ziefle, M., (2003). Design for all: User characteristics to be considered for the design of phones with hierarchical menu structures. In H. Luczak, K. J. Zink (Eds.), *Human factors in organizational design and management*, (pp. 503-508). Santa Monica: IEA.

Bay, S., & Ziefle, M. (2005). Children using cellular phones. The Effects of shortcomings in user interface design. *Human Factors*, *47*(1), 158-168.

Bay, S., & Ziefle, M. (in press). Landmarks or surveys? The impact of different instructions on children's performance in hierarchical menu structures. *Computers in Human Behavior*, (2007), doi:10.1016/j.chb.2007.05.003

Beckwith, L., Kyssinger, C., Burnett, M., Wiedenbeck, S., Lawrence, J., Blackwell, A., & Cook, C. (2006). Tinkering and gender in end-user programmers debbuging. In *Proceedings of the ACM Conference on Human Factors in Computing Systems 2006* (pp. 231-240).

Busch, T. (1995). Gender differences in self-efficacy and attitudes toward computers. *Journal of Educational Computing Research*, *12*, 147-158.

Davies, S. (1994). Knowledge restructuring and the acquisition of programming expertise. *International Journal of Human-Computer Studies, 40*, 703-726.

Downing, R.W., Moore, J.L., & Brown, S.W. (2005). The effects and interaction of spatial visualization and domain expertise on information seeking. *Computers in Human Behavior, 21*, 195-209.

Egan, D. (1988). Individual differences in human-computer-interaction. In M. Helander (Ed.), *Handbook of human-computer-interaction* (pp. 543-568). Amsterdam: Elsevier.

Ekstrom, R. B., French, J. W., Harman, H. H., & Dermen, D. (1976). *Manual for the kit of factor-referenced cognitive tests*. Princeton, NJ: Educational Testing Service.

Goodman, J., Gray, P., Khammampad, K., & Brewster, S. (2004) Using landmarks to support older people in navigation. In S. Brewster & M. Dunlop (Eds.), *Mobile human computer interaction* (pp. 38-48). Berlin, Germany: Springer.

Hasher, L., & Zacks, R. (1988). Working memory, comprehension, and aging: A review and a new view. In G.H. Bower (Ed.), *The psychology of learning and motivation* (pp. 193-225). San Diego: Academic.

Kim, H., & Hirtle, S. (1995). Spatial metaphors and disorientation in hypertext browsing. *Behaviour & Information Technology, 14*, 239-250.

Jakobs, E.-M. (2005). Technikakzeptanz und Technikteilhabe [acceptance and usage of Technology]. *Technikfolgenabschätzung*, *14*(3), 68-75.

Maguire, M., & Osman, Z. (2003). Designing for older and inexperienced mobile phone users. In C. Stephanidis (Ed.), *Universal access in HCI* (pp. 439-443). Mahwah, NJ: LEA.

Mobileyouth.org, 5 Million School Children Own A Mobile Phone in UK. Available online at: http://www.mobileyouth.org/my_item.php?mid=1471 (accessed 2005).

Rodger, J. A., & Pendharkar, P. C. (2004). A field study of the impact of gender and user's technical experience on the performance of voice-activated medical tracking application. *International Journal of Human-Computer Studies, 60*, 529-544.

Sein, M. K., Olfman, L., Bostrom, R. P., & Davies, S.A. (1993). Visualisation ability as a prediction of user learning success. *International Journal of Man-Machine-Studies*, *39*, 599-620.

Shemakin, F. N. (1962). Orientation in space. In B.G. Anan'yev et al. (Eds.). *Psychological sciences in the USSR*. (Vol. 1, Part 1) (Rep. No. 11466) (pp. 379-384). Washington: U.S. Office of Technical Reports.

Tuomainen, K., & Haapanen, S. (2003). Needs of the active elderly for mobile phones. In C. Stephanidis (Ed.), *Universal access in HCI: Inclusive design in the information society* (pp. 494-498). Mahwah, NJ: LEA.

Van-den Heuvel-Panheizen, M. (1999). Girls' and boys' problems: gender differences in solving problems in primary school mathematics in the Netherlands. In T. Nunes and P. Bryant (Eds.), *Learning and teaching mathematics* (pp. 223-253). Psychology Press.

Vicente, K.J., Hayes, B.C., & Williges, R.C. (1987). Assaying and isolating individual differences in searching a hierarchical files system. *Human Factors, 29(3)*, 349-359.

Westerman, S.J. (1997). Individual differences in the use of command line and menu computer interfaces. *International Journal of Human Computer Interaction*, *9(2)*, 183-198.

Ziefle, M., & Bay, S. (2004). Mental models of a cellular phone menu. Comparing older and younger novice. In S. Brewster & M. Dunlop (Eds.), *Mobile human computer interaction* (pp. 25-37). Berlin, Germany: Springer.

Ziefle, M., & Bay, S. (2005). How older adults meet complexity: Ageing effects on the usability of different mobile phones. *Behaviour & Information Technology, 24*(5), 375-389.

Ziefle, M., Arning, K., & Bay, S. (2006). Cross platform consistency and cognitive compatibility: the importance of users' mental model for the interaction with mobile devices. In K. Richter, J. Nichols, K. Gajos & A. Seffah (Eds.), *MAFOC' 06. The Many Faces of Consistency in Cross-Platform Design* (pp. 75-81). Retrieved in 2007 from http://CEUR-WS.org/Vol.198

Ziefle, M., Bay, S., & Schwade, A. (2006). On keys' meanings and modes: The impact of navigation key solutions on children's efficiency using a mobile phone. *Behaviour & Information Technology, 25*(5), 413-434.

KEY TERMS

Ease of Use: The ease of use describes the extent to which users believe a technical system to be free from effort and easy to handle.

Effectiveness: The term reflects the degree to which system objectives (e.g., tasks) are being achieved.

Efficiency: The term describes the degree to which a certain performance is achieved in terms of productivity. For example, it can be analyzed how many detouring routes are carried out until users reach their targeted goal in the menu, and how often they return to higher levels in menu hierarchy to re-orientate.

Navigation Performance: The term navigation in this context describes the process of moving through a menu structure in order to retrieve information or choose functions. The individual navigation routes the users take while searching for a specific target function may give valuable insights into shortcomings of menu design.

Trans-Generational Designs: Trans-generational designs are interface designs, which are usable and understandable by a broad user group, thus meeting the needs and demands of user diversity. They aim at coming up with developmental (cognitive, physical and sensory) specificities, which are present in users of different ages.

Usability: The term describes users' effectiveness, efficiency, and satisfaction with which users achieve specified goals in a technical system.

Spatial Ability: It is conceptualized as the ability to mentally manipulate and integrate visual stimuli consisting of more than one part. This includes the ability to imagine of rotations of objects or their parts.

Verbal Memory: Verbal memory is the basic ability to store and retrieve verbal or semantic information without additional processing.

Chapter IX
Learning-Disabled Children:
A Disregarded User Group

Susanne Bay
RWTH Aachen University, Germany

Martina Ziefle
RWTH Aachen University, Germany

ABSTRACT

In usability research it is a common practice to take young and healthy university students as participants for usability evaluations. This chapter focuses on the "weaker" mobile phone users, which have been mostly disregarded in this field: Learning-disabled children. Their interaction with mobile phones is compared to that of average children and students. Results show that the consideration of the "ergonomic worst case," which means a user group with cognitive deficiencies, leads to qualitatively and quantitatively different insights into the impact of specific design decisions. In contrast, when only students are involved as participants in the evaluation of technical devices, the impact of characteristics of the user interface on the ease of use is dramatically underestimated. One factor hampering the ability of learning-disabled children to interact meaningfully with a technical device may be their big difficulty building a correct mental representation of it. Therefore, this process should be especially supported.

INTRODUCTION

In most research projects, focusing on the usability of technical devices students serve as participants for the experimental evaluations. As students are bright and technically skilled, highly performance-motivated, have high cognitive and verbal abilities and no fear of being tested, the examination of this user group can be regarded as benchmark. Of course, this may give an insight into the effects of a specific design on users' performance interacting with the device because results can be interpreted as mainly caused by the design of the technical device, and no shortcomings have to be considered from the users' side. Furthermore, there are practical reasons for this procedure, as students can be recruited very easily by research institutions. On the other hand, the fact that some devices such as the mobile phone can be found in all age groups and levels of society give reason to seriously doubt whether students as participants in usability tests will be able to identify the real impact of specific user interface alternatives on the ease of use of the device. Taking only students' performance as basis for design decisions seems to be risky.

The purpose of the present study was to learn if and to which extent the performance achieved by students in usability studies may be generalized to a broader (or weaker) population. The motivation was to assess with a common technical device and typical tasks whether not only the quantitative performance level but also specific difficulties of the special user group could be identified. If the same difficulties may be found this means that the user interface design should be aligned with the "weakest" user's needs. If specific problems are encountered, a "design for all" approach would not be feasible but special design recommendations for different user groups would be needed.

BACKGROUND

Considering the variance in all factors characterizing the users it is highly debatable if the benchmark procedure for usability evaluations meets the demands of easily usable devices for all target users. There are differences, for example in expertise, experience with technology in general, domain knowledge, cultural factors and upbringing, but also developmental aspects with respect to the huge field of cognitive abilities, ascending in children and descending in older adults. As shown in earlier studies (e.g., Bay & Ziefle, 2003a; Liben, Patterson & Newcombe, 1981; Vicente, Hayes & Williges, 1987; Westermann, 1997; Ziefle & Bay, 2005a, Ziefle & Bay, 2006) a number of cognitive abilities, for example spatial cognition or verbal memory, show a considerable change over the lifespan.

Given that diversity, it may be problematic to focus only on best case conditions and to neglect weaker users. Rather, it might be more advisable to pursue an inverse proceeding in usability research in order to reach what usable designs promise.

Everyday, products as the mobile phone should be conceptualized bearing in mind the "weaker" user, that is, for example, a user with cognitive abilities below average. These users are the ones who need to be supported much more than those who are well trained with technical devices and office software because otherwise they may not be able to handle a device even after a substantial time of exposure. Also, more and more children possess mobile phones, which have not been specifically designed for this user group. Probably children would not even want to use a "kid's phone" because of "image" issues. In the recent past a number of studies have been concerned with enlightening children as a special user group of technical devices or technology in general (Berg, Taylor & Harper, 2003; Carusi & Mont'Alvao, 2006; Hanna et al., 1998; Jones & Liu, 1997; Ketola & Kohonen, 2001; Lieberman, 1998). While some knowledge was collected on children's attitudes (e.g., Vincent, 2004) and general usage criteria (e.g., Crenzel & Nojima, 2006), only few studies have investigated *how* children actually interact with different mobile phones in terms of efficiency and effectiveness (e.g., Bay & Ziefle, 2003b; Bay & Ziefle, 2005; Ziefle, Bay & Schwade, 2006). And in even fewer studies a direct comparison of the children with the performance of other user groups (i.e., young and older adults) was undertaken (Ziefle & Bay, 2004; 2005b).

Similar to the small HCI research output regarding children there is even less knowledge about mentally impaired users' interaction with technology (e.g., Oliver et al., 2001; Petrie et al., 2006; Mátrai, Kosztyán & Sik-Lányi, in press). Especially for these people the importance of usable mobile devices is high. Given the fact that mere calling is not longer the most frequent interaction but impaired users could be supported by memory functionality (e.g., medical monitoring) or navigation aids of mobiles, the mobile device could be a supportive aid enabling more independency and higher mobility of this special group.

Learning-disabled children are said to show the same developmental process as "normal" children, their developmental speed is only slower (on average about 1 to 2 years behind average children). Usually, they do not reach the highest stadium of cognitive development, which is characterized by abstract reasoning about problems (Schröder, 2000). Learning-disabled show a permanently constricted learning field, which means they are only susceptive to concrete and needs-related material. They have a reduced ability for abstractions, limited capacity to structure tasks and are generally slow, shallow and time-limited in their learning process.

In Germany, around 2.2 percent of all pupils are characterized as learning-disabled and need therefore special education. They are a marginal group but they deserve some attention from ergonomists and designers, because they are currently totally disregarded in usability research. Furthermore, a detailed look at their difficulties interacting with a mobile phone can give interesting insights into problems an average user will also very likely experience, for example when his attention is not entirely focused on the phone because he is on the move. When pursuing a "design for all" approach, learning-disabled are certainly a prototypical user group that should be considered for participation in usability studies.

MAIN FOCUS OF THE CHAPTER

In the present study the performance of different user groups interacting with mobile phones is compared: A student group and two groups of children—one being average school kids between 9 and 12 years of age, the second consisting of learning-disabled children and teenagers between 11 and 15 years of age. Their performance when solving typical tasks on a widespread phone model, the Siemens C35i, is evaluated. To assess the impact of specific user interfaces of a mobile phone in different user groups, only one aspect of the phone is experimentally varied: the keys that are used to interact with the mobile phone's menu. For reasons of ecological validity the navigation keys of a second widespread mobile phone model, the Nokia 3210, were chosen. Only few adaptations of the menu (such as changing the position of the soft key labels on the display) were necessary to operate the menu of the Siemens C35i with the Nokia 3210 keys.

Method

Participants

In the experiment, three different user groups with a total of 80 participants took part. Thirty students, 20 children with normal intelligence and 30 learning-disabled children and teenagers. The 20 children with normal intelligence were between 9 and 12 years of age, the 30 learning-disabled between 11 and 15 years. (The differences between those two groups regarding number of participants and age were due to difficulties recruiting these special users, yet the age difference is of minor importance since the "cognitive" age of learning-disabled is reduced compared to their "real" age. Half of the participants in each user group processed tasks using the phone with menu and keys both stemming from the Siemens C35i, the other half using the phone consisting of the Siemens C35i menu and the Nokia 3210 navigation keys.

Apparatus

The two mobile phones were simulated on a touch screen connected to a PC where user actions were logged on the keystroke level. To ensure good visibility and avoid difficulties hitting the keys because if the missing tactile feedback the display and the keys of the mobile phones were enlarged compared to the original devices. The appearance of the simulated phones was also modified to exclude effects of preferences for specific brands. The touch screen was fixated on a table in an angle of 35° which enabled an interaction in approximately the same posture as when using a real mobile phone.

Key Description

The two variants of navigation keys differed with regard to the number of keys (more specifically, the number of different options to be pressed) and the number of different functions each of these keys can exert. Moreover, among those keys which exert different functions at different points within the menu, there are such that have a similar meaning (e.g., confirm and save) and such that are quite dissimilar (e.g., end calls and return to a higher menu level).

The two original navigation key solutions which were simulated in the present experiment are shown in Figure 1 and 2.

The C35i keys consist in total of seven key options. Each of the two rocker switches contains two options (marked by a dot on each side of the rocker switch) were the key may be pressed, thus resulting in two "stroke options" per key. Sometimes (depending on the menu level) however, there are not two different options to be selected, but the same function is exerted, independently which side of the rocker switch that was pressed. The label displayed on the display above the key indicates which functions can be exerted. The left rocker switch has six different functions. The functions are scrolling (left: up, right: down), selecting the mailbox, changing, saving entries, and sometimes it has no function at all, depending on the point of the menu. All those actions are semantically very different from the scrolling function. The right rocker switch serves to enter the menu, to select, to correct (left part) and confirm (right part), or to correct (left) and save (right), and to send a message (eight functions/combinations of functions, where six are semantically dissimilar from selecting/confirming). Additionally, there is an extra key with an icon (open book) to open the phone directory. This function is most of the time not active at all. Furthermore, there is a big, centrally positioned key with a green receiver icon on it is used to make calls, which also exerts a function in specific cases (e.g., when a number is displayed), otherwise having no function. Finally, there is a smaller key with a red receiver sign to end calls as well as for hierarchical steps back in the menu.

Figure 1. Navigation keys of the Siemens C35i

Figure 2. Navigation keys of the Nokia 3210

The Nokia 3210 navigation keys exhibit four key options. Two of them have several functions: The c-key is used for corrections of letters and digits as well as for returns to higher menu levels. These two functions can be regarded as similar, as they both mean "undo." The centrally positioned key is a softkey used to enter the menu, to select highlighted menu entries, to confirm and to effect calls (four functions, three of them semantically similar representing confirmation actions, but entering the menu is not a confirmation action and can therefore be regarded as semantically dissimilar). The scrolling-key is used for movements up and down within any level of the menu.

Overall, the Nokia keys can be judged as simple with respect to both the number of keys and the number of keys with different functions compared to the Siemens keys.

Procedure

To assess the participants` previous experience using different kinds of technical devices, including the mobile phone, they were asked to complete a questionnaire before processing tasks on the mobile phone. The questionnaire was shown on the touch screen and required participants to activate fields by touching the screen. Thus, they were able to get used to the reaction of the touch logic. On a five-point scale the frequency of using different devices had to be answered (1= "several times a day," 2 = "once a day," 3 = "once a week," 4 = "once a month" and 5 = "less than once a month"). The estimated ease of use of those devices was to be judged on a four-point scale (1 = "very easy," 2 = "rather easy," 3 = "rather difficult," and 4 = "very difficult"). Also, the general interest in technology had to be rated on a four-point scale (1 = "low," 2 = "rather low," 3 = "rather strong," 4 = "strong").

Afterwards, participants had to complete four tasks on the simulated mobile phone:

1. Enter a telephone number and make a call
2. Send a text message (SMS) to a specific phone number (to compensate differences in the speed of typing the text, the message was provided and only had to be sent)
3. Hide your own phone number when calling
4. Redirect all phone calls to the mailbox

The participants were given a period of five minutes to solve each of the tasks. When a task was solved correctly, a "congratulations" message was shown on the screen. When a participant did not succeed in solving a task within the period of five minutes the experimenter told the participant that the specific tasks was very hard to solve (in order to prevent user's frustration) and that he should go on with the next.

Independent and Dependent Variables

The first independent variable was the key solution (3210 vs. C35i). The second independent variable was the user group (students vs. average children vs. learning-disabled children).

As dependent variable the participants' performance was assessed by counting the number of ineffective keystrokes carried out. That is, each key stroke that did not lead to any task related effect on the display. This includes:

- Hash (#) and asterisk (*) at any point within the menu
- Numbers when not task related
- Soft keys, function keys and scroll-buttons when not exerting a function

Additionally, the number of steps executed, the time needed to process tasks as well as the number of tasks solved were measured.

Results

The Participants' Experience with Technology

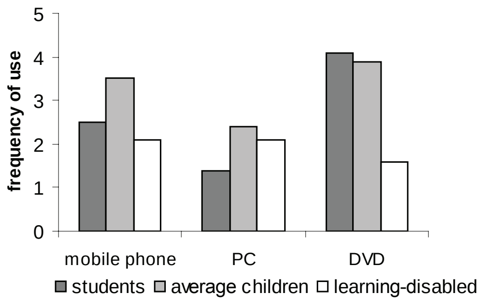
In a pre-experimental questionnaire the participants' experience with different technological devices was surveyed.

The students reported to use a mobile phone between daily and once a week (M = 2.5; SD = 1.6), the wireless phone once a week (M = 3.0; SD = 1.8), the PC between several times and once a day (M = 1.4; SD = 0, 7) and a DVD or VCR once a month (M = 4.1; SD = 0.9). The perceived ease of use of all devices was between "very easy" and "rather easy" (mobile phone: M = 1.7, SD = 0.9; wireless phone: M = 1.4, SD = 0.7; PC: M = 1.8, SD = 0.8; DVD: M = 1.7, SD = 0.7). On average, the students' interest in technology was rated between "rather strong" and "rather low" (M = 2.4, SD = 0.9).

The learning-disabled children had quite some experience using technical devices: The mobile phone (M = 2.1; SD = 1.6) and a PC (M = 2.2; SD = 0, 9) were used daily, the wireless phone (M = 3, 0; SD = 1.6) and DVD player (M = 2.9; SD = 1.3) once a week. The estimated ease of use of all devices was between "very easy" and "rather easy" (mobile phone: M = 1.5; SD = 0.8; PC: M = 1.7, SD = 0.9; wireless phone: M = 1.3, SD = 0.8; DVD: M = 1.6, SD = 1.1). The general interest of the learning-disabled in technology was rated as "rather high" (M = 3.3; SD = 1.1).

Looking at the average children's answers in the questionnaire it may be said that they were somewhat less experienced than the learning-disabled. They reported to use a mobile phone between once a week and once a month (M = 3.5; SD = 1.4), the PC between once a day and once a month (M = 2.4; SD = 1.1), the wireless phone

Figure 3. Frequency of using a mobile phone, a PC and a DVD player in the three user groups (1 = "several times a day," 2 = "once a day," 3 = "once a week," 4 = "once a month," 5 = "less often")



once a day (M = 2.0; SD = 0.9) and the DVD once a month (M = 3.9; SD = 1.0). The reported ease of use was between "very easy" and "rather easy" as in the two other user groups (mobile phone: M = 1.8, SD = 0.9; wireless phone: M = 1.1, SD = 0.3; PC: M = 1.8, SD = 0.8; DVD: M = 1.8, SD = 0.9). Children's general interest in technology was rather low (M = 2.1; SD = 1.0).

As visualized in Figure 3, it has to be stated that the learning-disabled are not less experienced with technology. Quite the contrary, their reported frequency with which they use a mobile phone and a DVD-player is even higher than that reported by the other two groups. Therefore, from this perspective no big performance differences should be expected.

In order to draw back performance differences between the participants using the different key solutions to the experimental manipulation, we had to make sure that the groups did not differ regarding their experience with technology. Therefore non-parametric Mann-Whitney tests were carried out for the variables surveyed in the questionnaire. No significant differences could be detected.

Performance Using the Two Phones

For each of the three groups of participants the number of ineffective keystrokes carried out with the two navigation key solutions is assessed.

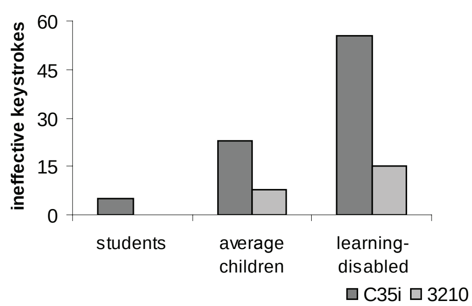
First the total number of ineffective keystrokes carried out by the students, the average children and the learning-disabled using the two phones is analyzed. Figure 4 shows the outcomes.

The students carried out 5.1 (SD = 6.3) ineffective keystrokes when using the C35i keys and 0.2 (SD = 0.6) with the 3210 key solution. Children in the C35i group made 22.8 (SD = 31.5) ineffective keystrokes, those of the 3210 group only 7.7 (SD = 7.7). Learning-disabled made 10 times as many ineffective keystrokes than the students and more than twice as many as average children when using the C35i (M = 55.3, SD = 71.2). Learning-disabled using the 3210 key solution made 14.9 (SD = 16.1) ineffective keystrokes, which also represents a considerably higher number compared to students and children. However, the performance difference between the two groups using different key solutions becomes more obvious in learning-disabled children (55.3 vs. 14.9 ineffective keystrokes) than in other user groups. Thus, the huge impact of different key solutions on user's performance becomes only apparent when participants other than students are taken into consideration. For a deeper insight into the performance of the different groups, the number of keystrokes carried out in each of the tasks is looked at in detail.

Effects of Key Solutions on Students

Figure 5 shows the performance outcomes for the first user group, the students, where the number of ineffective key strokes in each of the four tasks

Figure 4. Ineffective keystrokes carried out in the four tasks by the three user groups when using the Siemens C35i keys and the Nokia 3210 keys



is visualized. Students using the mobile phone with the Nokia 3210 keys carried out almost no ineffective keystroke (on average only 0.1 in tasks 2 (SD = 0.5) and 3 (SD = 0.3)) and also using the Siemens C35i keys the number of ineffective keystrokes may be regarded as negligible: When effecting a call no ineffective keystroke was carried out. When sending an SMS 2.9 (SD = 5.6), when hiding their own number 1.9 (SD = 3.8) and when making a call divert to the mailbox only 0.3 (SD = 0.7) times keys were stroked ineffectively. For the tasks of sending a short message ($t(28) = 1.87$; $p < .1$) and of hiding their own number ($t(28) = 1.91$; $p < .1$) marginally significant differences, depending on the key solution used, could be detected. Nevertheless, due to the small total number of ineffective keystrokes affected, it could be concluded, that the different key solutions do not significantly affect the performance interacting with the mobile phone.

But, before drawing this conclusion, a closer look should be taken at the performance of the other user groups.

Effects of Key Solutions on Average Children

Figure 6 visualizes performance outcomes for the average children carrying out the four tasks with different phones. It becomes evident that this user group undertakes considerably more keystrokes and also the difference between the two key solutions is somewhat clearer.

When effecting a call, children using the 3210 keys make 0.3 (SD = 0.9) ineffective keystrokes. With the C35i keys 0.6 (SD =2.0) ineffective keystrokes are undertaken. Thus, this task seems not to impose high demands on the children. In task two, where a text message was to be sent, with the 3210 keys 3.7 (SD = 6.8) and with the C35i 8.9 (SD = 10.0) ineffective keystrokes are undertaken. In the task of hiding their own number, participants press 2.8 (SD = 4.6) keys ineffectively when using the 3210, and 5.8 (SD =8.7) keys with the C35i solution. The last task, in which a call divert had to be carried out, led to the greatest performance difference between the two key solutions. Using the 3210, keys were stroked without exerting an effect only 0.7 (SD = 1.1) times, whereas in case of the C35i keys this happened on average 7.8 (SD = 15.2) times. In spite of the big numerical differences in performance between the users of the two phones, t-tests did not reveal any significant effect for any task, which may be due to the big variance in the data.

From these results obtained with children possessing average intelligence, it may be assumed that the navigation key solution exerts some effect on the users' performance interacting with a mobile phone, but the impact does not seem to be dramatic. After all, when solving tasks using the Siemens C35i on average only 9 times a key is stroked without exerting any effect. This should not unsettle a user. However, the results give a

Figure 5. Number of ineffective key strokes carried out by students using the two navigation key solutions

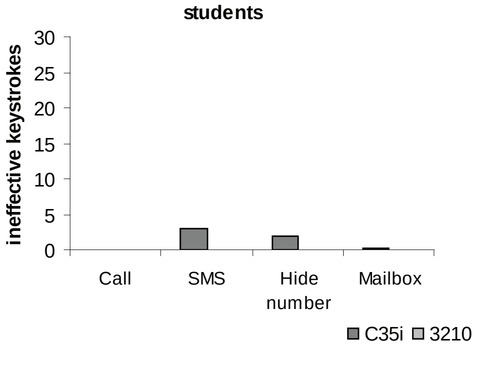
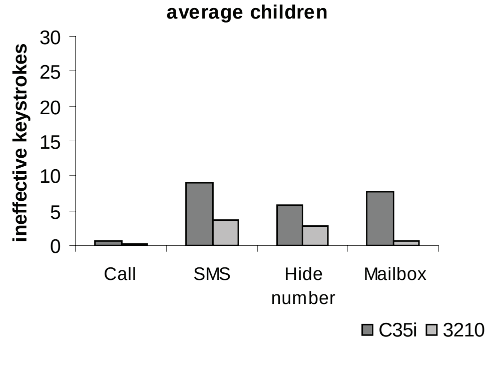


Figure 6. Number of ineffective keystrokes carried out by average children using the two navigation key solutions



first hint that different navigation key solutions used to control the same menu can lead to some interesting differences in performance.

For a deeper insight into the effects, a user group, which should be more sensitive towards complex rules of interaction, is surveyed.

Effects of Key Solutions on Learning-Disabled Children

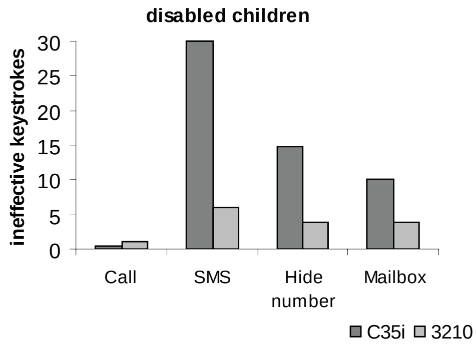
Figure 7 shows the number of ineffective keystrokes carried out by the group of learning-disabled children when using the Nokia 3210 keys and when using the original keys of the Siemens C35i.

Huge performance differences become evident. When using the 3210 keys between one and six ineffective keystrokes were carried out by the learning-disabled in the four tasks (call: M = 1.1, SD = 1.7; SMS: M = 6.1, SD = 7.4; hide own number: M = 3.8, SD = 5.7 and call divert to mailbox: M = 3.9, SD = 5.0). In contrast, when the learning-disabled children used the original keys of the C35i phone, averages of up to 30 ineffective keystrokes—when sending a text message, SD = 48.5—were reached. And also when hiding their own number and diverting calls to the mailbox with 14.8 (SD = 16.4) and 10.1 (SD = 15.3) a substantial number of ineffective keystrokes was undertaken. In the task of calling a number only 0.4 (SD = 1.1) ineffective keystrokes were made (Figure 7). T-test show significant differences between the two key solutions in the task of hiding their own number ($t(28) = 2.46$; $p < 0.05$) and a marginally significant difference in sending a text message ($t(28) = 1.89$; $p < 0.1$). This impressive performance differences between the two key solutions when they are used by learning-disabled should not be ignored. Interestingly, it is not the original key solution that leads to the best performance, but the solutions originating from an alternative phone, the Nokia 3210.

Which Keys Lead to the Difficulties?

The question, why the Siemens C35i keys led to many more ineffective keystrokes, may be answered by looking at the specific keys, which were stroked very often without exerting a function. The key that led to the biggest number of ineffective strokes was the receiver key, which is used to effect calls in the C35i solution. This key was used on average 28.7 times by the learning-disabled, 4.9 times by average children and on average 0.6 times by the students without exerting any effect. The key is comparably large, green and centrally positioned on the mobile phone, and was therefore presumably mistaken as a confirmation key.

Figure 7. Number of ineffective keystrokes carried out by learning-disabled children using the two navigation key solutions



The second type of keys, which were often used without exerting a function at the current point within the menu, were the soft keys of the Siemens C35i. Learning-disabled pressed these keys 17.7 times ineffectively, average children 9.2 times and students 4.1 times. The two soft keys of the C35i model exert many different functions at different points within the menu, and can sometimes exert different functions when stroked left or right. Sometimes, they have only one or even no functionality at all. This changing assignment of modes-of-operation has probably confused the users and especially the learning-disabled children, which led to many ineffective keystrokes.

Performance Differences between the User Groups in other Variables

For an insight into the performance differences between the three user groups of the present study two other variables, processing time and detour steps, are looked at. These variables are not directly related to the usability of the key solution, but are caused mainly by the difficulty imposed by the menu of the mobile phone. Therefore no differentiation between the two navigation key solutions used is made.

When effecting a call, both groups of children needed more than double the time (average children M = 44.2s; learning-disabled: M = 41.3s) than the students (M = 19.1s). In the task of sending a text message the differences between the groups increased: Students needed 83.1s, children 161.8 and learning-disabled 212.3s. Similar patterns of results were found for the tasks of hiding their own number and making a call divert to the mailbox. To hide their own number students needed 116.8 s, average children 221.4s and learning-disabled 242.7s. To make a call divert to the mailbox students needed 109.1s, average children 204.3s and learning-disabled 225.7s (Figure 8).

Thus, when considering only student participants, the difficulty imposed by the phone with a maximum of two minutes for completing the task of hiding their own number is not irrelevant, but still limited. However, when considering the performance of learning-disabled, who needed on average four minutes for this task, many were not able to actually solve it. The need for an improvement of the mobile phone's user interface becomes obvious.

A look at the number of steps executed while trying to solve the tasks confirms the argument. In the first task, where a number had to be entered and a call effected, performance between the groups does not differ meaningfully. Students need 12.9 steps, average children 16.9 and learning-disabled 17.4 steps. In the other tasks, learning-disabled children need mostly more than twice as many steps as the students. To send an SMS students need 51.2 steps, average children 68.6 and learning-disabled 120 steps. In the task of hiding their own number students execute 86.7 steps, average children 129.9 and learning-disabled 163.4 steps. In the last task students made 76.9 steps during their attempts to solve the task, average children made 141.4 and learning-disabled made 155.7 steps.

It becomes evident that the difficulty to perform different tasks on a widespread mobile phone is underestimated when only students are selected as participants for usability tests.

Figure 8. Time needed to process the four tasks by the three user groups

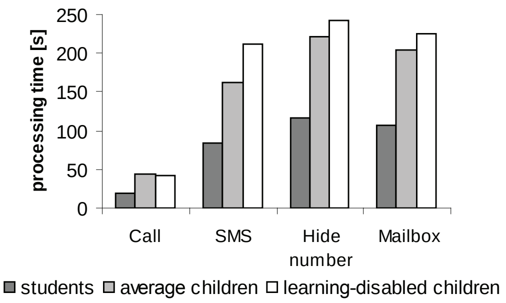
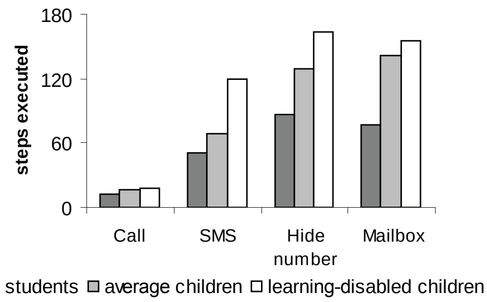


Figure 9. Steps executed while processing the four tasks by the three user groups



What is so Special About Learning-Disabled?

The results outlined above give reason to argue that designers of interfaces for mobile phones (and presumably of other electronic devices, too) should focus on weaker mobile phone users, such as learning-disabled, if they want to make sure that their device is really usable for a broad range of users. However, to understand what makes the learning-disabled a special group, it is worthwhile to examine their specific characteristics that may be of importance for the interaction with mobile devices.

Memory

The first important aspect, which differentiates learning-disabled from other users, is their memory capacity. The high correlation between memory capacity and the performance using a mobile phone was shown in a study by Bay and Ziefle (2003a). Participants of this study who were between 27 and 61 years of age had reached 12.6 points (SD = 2.2) in a test of figural memory. The learning-disabled of the present study were asked to solve the same memory test (LGT-3, Bäumler, 1974) and it revealed that the memory ability of the learning-disabled was somewhat lower with a mean of 9.8 (SD = 2.2) out of 20 points. This may be one reason for the big difficulties experienced by the learning-disabled. They may have had troubles remembering the functions, each of the different keys exerted, and which of the menu functions they had already selected. However, correlations between the scores in the memory test and performance measures, when using the mobile phone, did not reach the significance level in the present study.

Locus of Control

In earlier studies (Bay & Ziefle, 2003a; Ziefle, Bay & Schwade, 2006) it was found, that users' "experienced competency" with respect to the use of technical devices in general can also affect performance outcomes when using mobile phones. The locus of control regarding the use of technology (LOC) as measured through a standardized test (Beier, 1999) showed significant correlations with the number of tasks successfully solved on a mobile phone in a group of younger and older adults. Thus, users having high values in LOC showed a better performance than those with lower values. Even if it is not clear, whether the LOC is the antecedent of a good performance or if the successful interaction with technology is the antecedent of the high LOC value, the correlation shows that the felt competency and the real competency of using technical devices go in parallel in adults.

It is a basic question whether the learning-disabled have equally high self-assessment. On the one hand, it may be assumed that learning-disabled are intimidated by technical devices. This may be due to their negative experience interacting with them. Such an attitude could have negative impact on the way they approach technical problems and is therefore worth looking at. On the other hand, it is also equally plausible that the learning-disabled have no valid or realistic self-estimation with respect to their own competencies when using technical devices.

The learning-disabled children's locus of control interacting with technology was therefore surveyed with a standardized instrument by Beier (1999). It consists of eight statements, such as "I like cracking technical problems" or "Whenever I solve a technical problem this happens mostly by chance." These statements have to be affirmed or denied by the participants on a six point scale.

The learning-disabled reached an average score of 73.4 (SD = 15.7) of a maximum of 100 points. The participants in an earlier study (Bay & Ziefle, 2003a) had reached 66.1 points. The average children, in comparison, revealed a similar level with, on average, 67.4 points (SD = 14.1). The students showed a somewhat higher LOC level with 76.6 points (SD = 10.5), however, surprisingly, they did not reach values close to the upper end of the scale, as one could have expected.

Does the LOC really affect performance and can this be found in all user groups?

To assess interrelations between performance using the mobile phone and locus of control regarding technical devices, Spearman rank correlations were carried out. To begin with the benchmark, students showed a weak, but marginally significant correlation between the number of tasks solved and LOC values ($r = -0.3$; $p < 0.1$). Also, for the average children group LOC values were found to considerably affect performance (number of keys used ineffectively: $r = -0.47$; $p < 0.05$; tasks solved: $r = -0.50$; $p < 0.05$; time on task: $r = -0.36$; $p < 0.05$). Thus, for younger and older adults as well as for average children the LOC was interrelated with performance, even though to a different extent.

It characterizes the specificity of the disabled children that a significant correlation between LOC values and performance using the mobile phone could not be found in this group, for none of the dependent variables. This shows that this user group was not able to self-assess themselves realistically with respect to their technical competency. On the one hand, their absolute level in LOC was not very different from the other user groups, however, their performance level was distinctly lower: Their efficiency was 10 times lower with respect to the number of ineffectively used keys and the children needed double the time and twice as many detour steps when compared to the benchmark, the students' performance.

Mental Models

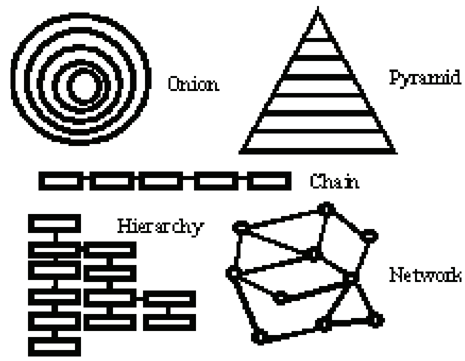
The importance of mental models for a purposeful interaction with technical devices has been emphasized by a number of studies (e.g., Norman, 1983), also more specifically for the interaction with mobile phones (Bay & Ziefle, 2003b; Ziefle & Bay, 2005). It was found that the better the mental representation of the spatial structure of the device (that is, the hierarchical nature in the case of mobile phones), the better was the performance of a user.

The difficulties of the learning-disabled children interacting with the mobile phones may therefore be due to a deficiency in building an appropriate mental representation of the menu structure. Therefore the users' mental model of the menu was assessed by showing the children a number of drawings that are supposed to visualize different kinds of mental models. The children had to process the phone tasks first, and were then asked to choose the one of the shown alternatives that was most appropriate, according to them. The different drawings of the mental models are shown in Figure 10.

The menu of a mobile phone has a hierarchical structure. However, only seven of the 30 learning-disabled children chose this drawing.

It is of interest, whether users who chose the correct drawing also performed better than those without a correct mental representation of the mobile phone menu. And indeed the analyses revealed a somewhat superior performance of the learning-disabled who were aware of the hierarchical nature of the menu. They executed 394.3

Figure 10. Drawings used to assess the users' mental representation of the menu structure

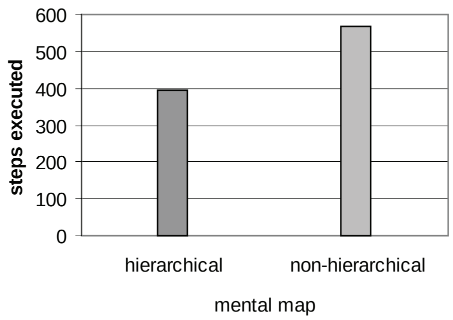


steps (SD = 148.2) in contrast to 569.6 steps (SD = 255.6) needed by the rest (Figure 11). This difference was marginally significant ($t(28) = 1.73$; $p < .1$). Users with a hierarchical mental map also solved somewhat more tasks (M = 2.9; SD = 1.1) and executed less ineffective keystrokes (M = 21.6; SD = 18.1) compared to the majority without a correct mental map who solved only 2.5 tasks (SD = 1.1) and made nearly twice as many ineffective keystrokes (M = 39.2; SD = 61.5). With regard to the processing time, differences were much smaller, but still showing a benefit of the correct mental representation (M = 829.4, SD = 251 versus M = 853.1, SD = 251.5 of the majority without a correct mental map).

Thus, the impact of a correct mental representation of the mobile phone menu on learning-disabled children's performance could be shown.

In a study by Bay and Ziefle (2003b), the benefit of having a correct model on the performance interacting with mobile phones was also found for average children. It was shown that nearly all (80 percent) of the examined children aged nine to 16 had a correct mental representation of the hierarchical nature of the menu. The results of the present study confirm the importance of a hierarchical mental representation of the menu and suggest that this is a crucial factor, which may explain the huge inferiority of the learning-disabled compared to other user groups including other children.

Figure 11. Performance difference between users with a hierarchical and a non-hierarchical mental representation of the menu



CONCLUSION

A study was conducted with the aim to explore the differences in performance between learning-disabled children, average children and students when interacting with a mobile phone. This study was undertaken out of two reasons: Firstly, learning-disabled have been completely disregarded as users of everyday technical products as the mobile phone. Secondly, the consideration of participants in usability research who do not possess high cognitive abilities can provide meaningful insights into the real implications of specific user interface design decisions. As one important example for a design decision the navigation key solution was varied. Participants processed four typical phone applications on two mobile phones simulated on a touch screen. One phone corresponded to the Siemens C35i regarding menu and navigation keys, the second phone had the same menu, which was to be operated with navigation keys stemming from the Nokia 3210 model.

Thirty students, 20 average children aged between 9 and 12 years of age and 30 learning-disabled children and teenagers between 11 and 15 years took part in the study. As expected, results revealed similarities between the three user groups as well as differences. The similarity between groups is the basic pattern of performance outcomes when using the C35i keys in contrast to the 3210 keys: All user groups showed a considerably better performance when using the simple and easy to understand 3210 keys compared to the rather complex C35i key solution. This means, that there are design solutions that benefit every user independent of age and cognitive abilities. This result makes the idea of the "design for all" approach feasible.

The difference that showed up between the user groups is the huge inferiority of the learning-disabled compared to the student group as well as to the average children. Learning-disabled students needed more than double the time and steps to process the tasks than the students and about one third more time and steps than average children for example when sending a text message. Also, the effect of the different key solutions becomes more evident in the learning-disabled compared to the other groups, which can be taken from the total amount of ineffectively used keys. In total, using the C35i keys the learning-disabled made on average 55 ineffective keystrokes while solving the four tasks, and this is actually ten times more than students and more than twice as many as the average children carried out when using the C35i keys. However, learning-disabled children made only 15 ineffective keystrokes with the 3210 keys, which shows how much this user group benefits from good design. With student users, this difference between the two key solutions did not become as obvious, since with the more complex phone only five ineffective keystrokes were carried out on average—an number that apparently may be neglected. Thus, the real difficulties that can be caused by a navigation key solution and the importance of creating less complex solutions only becomes evident when a user group like the learning-disabled is considered. If only students are taken as participants in usability tests—as it is very often the case—important aspects are completely ignored. For example, the present study could show how dramatic the impact of inconsistent assignment of functions to keys can be, which only reached a meaningful effect on performance in the group of learning-disabled users. Average children and students experienced the same difficulties but were able to overcome them after some practice.

It may be concluded that whenever a "design for all" approach is pursued, the ergonomic "worst case" has to be taken into consideration. Even if the recruitment of those users requires more time, it is worth the effort and represents a very insightful experience. But what is so special about the learning–disabled as user group?

As a first hypothesis it may be assumed that learning-disabled children and teenagers have lower expertise and experience using different technical devices. Results show, however, that the contrary is the case. The learning-disabled of the present study showed a higher frequency of using a mobile phone, a DVD-player and a PC than average children and even partly outreached the students. Thus, this hypothesis has to be rejected.

When memory ability, which was found to influence the ability to interact with technical devices, was assessed some, but again, no meaningful differences between learning-disabled and other user groups were detected. However, for this user group, correlations between memory ability and performance measures did not lead to significant interrelations.

Furthermore, the locus of control of the learning-disabled children did not show to differ meaningfully from that of other users. In contrast, the learning-disabled children and teenagers surveyed in the present study seemed to be pretty convinced of their ability to handle technology well (nearly as convinced as the average children and the students were). Not only their locus of control was high but also the estimated ease using a mobile phone, a PC or a DVD player was high: They were rated between "very easy" and "rather easy" to use by the participants. Thus, the reported expertise has also to be ruled out as explanatory variable.

A last aspect that may account for the inferior performance of learning-disabled children and teenagers is the lack of an appropriate mental model of the functioning of the mobile phone, more specifically of the menu structure. And indeed only seven out of 30 learning-disabled were able to identify the hierarchy as the correct model of the menu structure. For a comparison, in Bay & Ziefle (2005) it was found that 80 percent of 9 to 16 years old children were aware of the hierarchical nature of the menu. Also, the present study showed that learning-disabled children possessing a correct mental map performed better in solving tasks on the mobile phone than participants whose mental map was incorrect.

It is often assumed that people automatically build-up cognitive representations of the functioning of technical devices while interacting with them (e.g., Norman, 1983). Thus, frequent exploration and active handling of the technical device is believed to support the development of adequate representations with respect to how the mental room, which has to be navigated, is structured. Furthermore, the exploration of the menu structure is assumed to pre-structure the interconnections and relations between functions and sub-categories present in the menu. However, learning-disabled apparently have big difficulties building a correct representation of the menu structure, considering that they reported to use a mobile phone on a daily basis. At least, the mental representation they have built does not correspond to the real information structure of the phone. Therefore, it may be concluded that the process of building a mental representation needs to be actively supported. This may be done through training or better visual cues on the spatial structure to be incorporated on the mobile phone's display (Ziefle & Bay, 2006).

Some final remarks are concerned with potential methodological limitations of the presented research study. One could critically argue that the phones under study were simulations on a touch screen rather than real mobile phones. This criticism can be met with three arguments: First, as only the keys were under study and have been experimentally varied independently of differences with regard to the menu, there were no real phones available that met these requirements. Second, different mobile phones differ in so many attributes and aspects (size, form, color, key shape, labels, haptics, etc.) that any comparison would be misleading for the question at issue, because we actually do not know which of the aspects leads to performance differences. Third, with the special user group under study—learning- disabled children—it was a deliberate aim to design the experimental situation as easy and comfortable as possible and with the simulation visual as well as psychomotor difficulties could be ruled out. It is clear that our results therefore represent an underestimation of the real performance but performance differences can be unequivocally traced back to the experimental variation. The same argument can serve to explain the feedback provided when a task was solved successfully. Whenever the children had solved a task they were presented a "Congratulations!" message—something that is not given in real life situations and definitively makes the experimental situation easier than in reality. However, as average as well as handicapped children use to quickly lose their motivation it was of central importance to sustain children's enthusiasm taking part in the experiment. Even though an underestimation of performance differences between the phones may have taken place, the key result that handicapped children show the very same structural problems with phones but react more sensitive on specific user interface design decisions in not affected by these methodological issues.

FUTURE TRENDS

Future studies should focus on supporting the process of cognitive mapping in participants. Even if the performance outcomes were qualitatively very similar between the different user groups, it may not necessarily be deduced that the development of a proper mental model is also similar. In the contrary, it is very likely that different user groups need different types of support for this purpose. Different types of training, instructions in manuals or information visualization on the display of the mobile phone itself should be evaluated with respect to their helpfulness for different user groups.

ACKNOWLEDGMENT

Thanks to Lisa Ansorge and Alexander Schwade for their help collecting and analyzing the data as well as Hans-Jürgen Bay for helpful comments on an earlier version of this chapter.

REFERENCES

Bäumler, G. (1974). *Lern- und Gedächtnistest 3 (LGT-3)* [Learning and Memory Test 3]. Göttingen, Germany: Hogrefe

Bay, S., & Ziefle, M. (2003a). Design for all: User characteristics to be considered for the design of devices with hierarchical menu structures. In H. Luczak & K.J. Zink (Eds.), *Human factors in organizational design and management* (pp. 503-508). Santa Monica: IEA.

Bay, S., & Ziefle, M. (2003b). Performance in mobile phones: Does it depend on proper cognitive mapping? In D. Harris; V. Duffy; M. Smith & C. Stephanidis (Eds), *Human centred computing. Cognitive, social and ergonomic aspects* (pp. 170-174). Mahwah, NJ: LEA.

Bay, S., & Ziefle, M. (2005). Children using cellular phones. The effects of shortcomings in user interface design. *Human Factors, 47*(1), 158-168.

Beier, G. (1999) Kontrollüberzeugung im Umgang mit Technik [Locus of control interacting with technology]. *Report Psychologie, 9*, 684-693.

Berg, S., Taylor, A., & Harper, R. (2003). Mobile phones for the next generation. Devices for Teenagers. *CHI, 5*(1), 433-440.

Carusi, A., & Mont`Alvao, C. (2006). Navigation in children's educational software: The influence of multimedia elements. In R. N. Pikaar; E.A. Konigsveld & P.J. Settels (Eds.), *Meeting diversity in ergonomics. Proceedings of the International Ergonomics Society 2006 Congress.* Amsterdam: Elsevier.

Crenzel, S.R., & Nojima. V.L. (2006). Children and instant messaging. In R. N. Pikaar; E.A. Konigsveld & P.J. Settels (Eds.), *Meeting diversity in ergonomics. Proceedings of the International Ergonomics Society 2006 Congress.* Amsterdam: Elsevier.

Hanna, L., Risden, K., Czerwinski, M., & Alexander, K.J. (1998). The role of usability research in designing children's computer products. In A. Druin (Ed.), *The design of children's technology*. San Francisco: Morgan Kaufmann Publishers.

Jones, M., & Liu, M. (1997). Introducing interactive multimedia to young children: A case study of how two-year-olds interact with the technology. *Journal of Computing in Childhood Education*, *8*(4), 313-343.

Ketola, P., & Korhonen, H. (2001). ToyMobile: Image-based telecommunication and small children. In: A. Blanford, J. Vanderdonckt & P. Gray (Eds.), *People and computers XV — interaction without frontiers, Proceedings of HCI 2001,* (pp. 415-426).

Liben, L. S., Patterson, A. H. & Newcombe, N. (Eds.), *Spatial representation and behavior across the life span* (pp. 195-233). New York: Academic Press.

Lieberman, D.A. (1998). The researcher's role in the design of children's media and technology. In A. Druin (Ed.), *The design of children's technology* (pp.73-97). San Francisco: Morgan Kaufmann Publishers.

Mátrai, R., Kosztyán, Z. T., & Sik-Lányi, C. (in press). *Navigation methods of special needs users in multimedia systems.* Computers in Human Behavior (2007), doi10.1016/j.chb.2007.07.015.

Norman, D. A. (1983). Some observations on mental models. In D. Gentner and A.L. Stevens (Eds.), *Mental models* (pp.7-14). Hillsdale, NJ: LEA.

Oliver, R., Gyi, D., Porter, M., Marshall, R. & Case, K. (2001). A survey of the design needs of older and disabled people. In M.A. Hanson (Ed.), *Contemporary ergonomics* (pp. 365-270). London, New York: Taylor & Francis.

Petrie, H., Hamilton, F., Fraser, N. & Pavan, P. (2006). Remote usability evaluations with disabled people. *In Proceedings of CHI* (pp. 1133-1141). Montreal: ACM.

Schröder, U. (2000) *Lernbehinderten Pädagogik, Grundlagen und Perspektiven sonderpädagogischer Lernhilfe.* [Pedagogics for learning-disabled, basics and perspectives of special education learning aid]. Stuttgart, Germany: Kohlhammer.

Vicente, K.J., Hayes, B.C., & Williges, R.C. (1987). Assaying and isolating individual differences in searching a hierarchical files system. *Human Factors, 29(3)*, 349-359.

Vincent, J. (2004). *'11 16 Mobile.' Examining mobile phone and ICT use amongst children aged 11 to 16*. Digital World Research Centre. Retrieved January 2007, from http://www.dwrc.surrey.ac.uk/Portals/0/11-16Mobiles.pdf

Westerman, S.J. (1997). Individual differences in the use of command line and menu computer interfaces. *International Journal of Human Computer Interaction, 9(2)*, 183-198.

Ziefle, M., & Bay, S. (2004). Mental models of a cellular phone menu. Comparing older and younger novice users. *Proceedings of Mobile HCI 2004, 6th International Symposium on Mobile Human-Computer Interaction* (pp. 25-37). Heidelberg, Germany: Springer.

Ziefle, M., & Bay, S. (2005a). How older adults meet cognitive complexity: Aging effects on the usability of different cellular phones. *Behaviour & Information Technology, 24*(5), 375-389.

Ziefle, M., & Bay, S. (2005b). The complexity of navigation keys in cellular phones and their effects on performance of teenagers and older adult phone users. In: *Proceedings of the HCI International 2005.(Vol. 4)* Theories, Models and processes in Human Computer Interaction. Mira Digital Publishing.

Ziefle, M., & Bay, S. (2006). How to overcome disorientation in mobile phone menus: A comparison of two different types of navigation aids. *Human Computer Interaction, 21*(4), 393-432.

Ziefle, M., Bay, S., & Schwade, A. (2006). On keys' meanings and modes: The impact of navigation key solutions on children's efficiency using a mobile phone. *Behaviour and Information Technology, 25*(5), 413-431.

KEY TERMS

Ecological Validity: Degree to which results of experiments are transferable to behavior in real world situations. The higher the ecological validity of an experiment the higher is the probability that results found in the experiment can be found in the same fashion in the field.

Ineffective Keystrokes: Measure for performance evaluation. Counting each key stroke carried out by a user that does not lead to any task related effect on the display enables to measure the difficulty imposed by the navigation keys of a mobile phone independent of the difficulties caused by the menu. Ineffective keystrokes include Hash (#) and asterisk (*) at any point within the menu, number keys when not task related as well as soft keys, function keys and scroll-buttons when not exerting a function.

Learning-Disabled Children: Children with a developmental speed that is slower than that of average children (about 1 to 2 years behind). Usually learning-disabled also do not reach the highest stadium of cognitive development which is characterized by abstract reasoning about problems. Learning-disabled show a permanently constricted learning field, which means they are only susceptive to concrete and needs-related material, they have a reduced ability for abstractions, limited capacity to structure tasks and are generally slow, shallow and time-limited in their learning process.

Mental Models: concepts in the mind of users about the functioning of devices, metaphors, and ideas which lead the user while interacting with the device.

Mobile Phone Menu: Form of displaying mobile phone functions that go beyond effectuation of calls to the user. Mobile phone menus usually have a hierarchical tree structure, which the user needs to navigate through via keys in order to find and select the desired function.

Navigation Keys: Keys used to operate the menu of a mobile phone, usually consisting at least of two keys fro scrolling up and down within one level, one key for selection and one key for returning to higher menu levels.

Chapter X
Human Factors Problems of Wearable Computers

Chris Baber
The University of Birmingham, UK

James Knight
The University of Birmingham, UK

ABSTRACT

In this chapter wearable computers are considered from the perspective of human factors. The basic argument is that wearable computers can be considered as a form of prosthesis. In broad terms, a prosthesis could be considered in terms of replacement (i.e., for damaged limbs or organs), correction (i.e., correction to 'normal' vision or hearing with glasses or hearing aids), or enhancement of some capability. Wearable computers offer the potential to enhance cognitive performance and as such could act as cognitive prosthesis, rather than as a physical prosthesis. However, wearable computers research is still very much at the stage of determining how the device is to be added to the body and what capability we are enhancing.

INTRODUCTION

There is a wide range of technologies that have been developed to be fitted to the person. Depending on one's definition of "technology," this could range from clothing and textiles, through to spectacles, to cochlear implants. The use of these different technologies can be basically summarized as the supplementation or augmentation of human capability, for example the ability to regulate core temperature (clothing), to see (spectacles) or to hear (cochlear implant). One reason why such supplementation might be required is that the current capability does not fit with environmental demands, either because the environment exceeds the limits over which the human body can function or because the capability is impaired or limited. From this perspective, a question for wearable computers should be what are the current human capabilities that are exceeded by the environment and require supplementation by wearable computers? In the majority of cases for wearable computers, the answer to this question hinges on communicative, perceptual, or cognitive ability. As Clark (2006) notes, "...the use, reach, and transformative powers of these cognitive technologies is escalating"(p. 2). but the essential point to note is that "Cognitive technologies are best understood as deep and integral parts of the problem-solving systems that constitute human intelligence"(p.2). Thus, such technologies could represent a form of 'cognitive prosthesis' in that they are intended to support cognitive activities, for example, having a camera (performing face-recognition) to advise the wearer on the name of the person in front of them. The immediate challenge is not necessarily one of technology but of cognition. If the technology is 'doing the recognition,' the question is raised what is the human left to do? To draw on a commonly cited analogy, spectacles serve as a perceptual prosthesis, that is, to improve or correct a person's vision. For wearable computers (and the related field of augmented reality), the 'improvement' could be to reveal to the person objects that are not present by overlaying an artificial display onto the world. The display could simply take the form of labels or directional arrows, or could be more a sophisticated presentation of moving (virtual) objects. In both cases, the 'augmentation' could either enhance the person's understanding of the environment or could substitute this understanding.

In terms of communication, mobile telephones and MP3 devices contain significant computing power such that they can easily be considered as being computers, albeit with limited functionality. These devices can be worn, for example MP3 players can be worn on the upper arm, attached to belts or neck-straps, or placed in hats, and mobile telephones can be attached to belts. Furthermore, with the wireless (Bluetooth) headset, the user interface of a mobile telephone can be worn on the ear at all times. Thus, both can be always present and both can be considered part of the person. A definition of wearable computers ought to, at least, allow differentiation from devices that can slip into the user's pockets (if this technology is to be treated as a new area of research and development). Two early definitions of wearable computers, from Bass (1996) and Mann (1997), emphasize that wearable computers are designed to exist within the corporeal envelope of the user and that this makes them part of what the user considers himself or herself. In many respects this allows an analogy to be drawn between wearable computers and prosthetic devices. Having something added to the body, whether externally, such as spectacles, artificial limbs, hearing aids, or internally, such as pace-makers, or cochlear implants, changes the performance of the person and (for external prosthesis) the appearance of the body. This now becomes a very different concept from the mobile telephone, MP3 player and the computer that we traditionally encounter. This raises all manner of interesting questions relating to physical, perceptual, and cognitive aspects of human factors, as well as a whole host of emotional aspects of wearing devices (for the wearer and the people with whom they interact). At the moment, there remains a gap between what a wearable computer is intended to be and what mobile telephones and MP3 players currently are. This gap can best be considered as a form of perceptual and cognitive prosthesis, in which the wearer's ability to view the world, retrieve pertinent information, and respond to environmental demands are enhanced by the technology across all aspects of everyday life. At present mobile telephones and MP3 players are able to be tailored (by the user) and can deal with a limited set of situations (relating to communications or music playing) but do not fill the specification that one might have for a wearable computer. The basic difference between a wearable computer and these other technologies lies in the question of how well the user can interact with both the device and the environment simultaneously. Obviously, listening to an MP3 player or speaking on a mobile telephone can be performed while walking through an environment. However, performing control actions on the devices can be sufficiently demanding to draw the user's attention from the environment. The ideal wearable computer would allow the user to manage attention to both device and environment.

In other words, it should allow the user to manage both foreground and background interaction (Hinckley et al., 2005).

FORM-FACTOR AND PHYSICAL ATTACHMENT

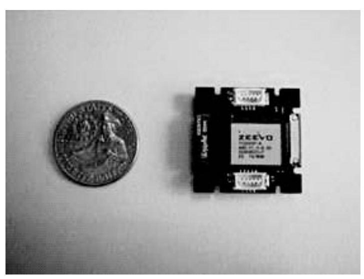
Moore's law continues to guarantee that the processors will get smaller, and work in the field of micro-electrical-mechanical systems (MEMS) shows how it is possible to create application specific processors that are small enough to be incorporated into buttons on clothing or into jewelry (Figure 1a).

Furthermore, it is feasible to assume the widespread development of general purpose processors, such as the mote (and similar) concept, that combine low power with sufficient processing capability to deal with a number of different sensors (figure 1b).

One direction for wearable computers is that the miniaturization of technology will mean it is possible to implant processors under the skin (or embed them into clothing). This is not a particularly novel idea as the medical world has been experimenting with mechanical implants for many years and has, over the past decade or so, developed digital implants that can, for example, regulate drug administration or improve hearing ability. While the development of such technology is exciting and likely to lead to fascinating discoveries, it lies somewhat outside the remit of this chapter. The problem is how can these extremely small devices support *interaction* with the person wearing (or supporting) them? To a great extent, the human simply has these devices fitted into them. While this represents the logical extension of the argument that the wearable compute exists within the wearers 'corporeal envelope,' it does preclude the possibility of developing means of interacting with the device. Indeed, as technology continues to get smaller, ever more difficult the challenge of supporting human-computer interaction becomes, for example, how can one read displays with very small font or press buttons that are much smaller than the human finger? This could mean that we should redefine our concept of *interaction,* for example, one could argue that 'interaction' should be any activity that the person performs, that is sensed by the device and that allows the device to make a response. A potential problem with such a concept is that the person might not be able to selectively control what the device senses or to fully understand why the device is behaving in the manner that it is. This means that one either locks the human out of the interaction (in which case one has a sensor system that acts on the environment and which coincidentally affects people in that environment) or one must develop different ways in which the person can understand and manage the behavior of the device. This is not a trivial problem and one can point to many developments in the ubiquitous and pervasive computing domain in which the role of the human in device performance is merely one of passive recipient of device activity. From the point of view of human-computer interaction, this will not only lead to

Figure 1a. MEMS device

Figure 1b. Intel's Mote prototype[1]



the ‘irony of automation’[2] but also to frustration, annoyance and a perceived lack of control. In other words, who wants to live in a world that is basically run by the descendants of the Microsoft Office Paperclip?

Physical Effects of Wearing Computers

We might expect a computer to consist of a processor and storage, some form of power, a display, and an interaction device. For wearable computers, each of these components can be separated and worn on different parts of the body. Thus, as the prototypes in Figure 2 illustrate, the processor and storage could be incorporated into a single unit and mounted on the waist or the back, the display could be mounted on the head (either near the eyes or the ears, depending on the feedback provided to the wearer), the power could be located near the processor unit, and the interaction device could be mounted within easy reach of the hand (if manual control) or the mouth (if speech input). These prototypes are typical of much contemporary work on wearable computers in terms of the relative size and placement of components.

Wearable computers represent a load on the person and consequently can affect the physical activity of the person (Zingale et al., 2005). Legg (1985) proposes that the human body can carry loads at the following points: head, shoulder, back, chest, trunk, upper arm, forearm, hands, thighs, feet, a combination of these, and aided (i.e., by pulling, pushing or sharing a load). A glance through the literature of wearable computers shows that most of these sites have been experimented with in various designs. Gemperle et al. (1998) offer the term ‘wearability’ to describe the use of the human body to physically support a given product and, by extension, the term ‘dynamic wearability’ to address the device being worn while the body is in motion. Given this notion of wearability, there is the question of where a device might be positioned on the body, and, once positioned, how it might affect the wearer in terms of balance, posture and musculoskeletal loading Given the notion of dynamic wearability, there are the questions of how the device will be carried or worn, how this might affect movement, and how this might lead to either perceptions of differences in movement patterns, physiological strain or psychological stress on the wearer. These changes can be assessed, using subjective self-report techniques (Bodine & Gemperle, 2003; Knight et al., 2002, 2006; Knight & Baber, 2005) and through objective analysis (Nigg & Herzog, 1994).

Figure 2. Wearing prototype wearable computers

Energy Expenditure, Muscle Activity and Using Wearable Computers

An attached load to the body will have a direct affect on the energy expended by the body, as the muscles burn more energy to generate force to counteract the weight of the load. This situation sees a wearer of a wearable computer potentially increasing their overall energy expenditure to overcome inertial changes. Thus, *any* additional weight on the body can create demands on the musculoskeletal system to support and move the object. In addition, the kinematics of posture and movement can compound the loading effects of a wearable computer. Knight and Baber (2004) report the following about head-mounted loads, such as head mounted displays: (1) different head postures have a measurable effect on musculo-skeletal loading; (2) the heavier the frontal load (as is representative of the loading of a HMD), the greater the muscle activity required to keep the head in a fixed position and; (3) in neutral or extended postures, the wearer can sustain larger loads than if the head is in a flexed or rotated head position Thus, it is not simply a matter of reducing load of head-mounted equipment, but one must determine the posture that the wearer of such equipment is likely to adopt. An alternative location for display technology is on the forearm. In their study of pointing devices, Thomas et al. (1998) demonstrate that for general purpose activity, the forearm would be an appropriate place to mount the device. This location is attractive as it allows the wearer to easily move the display into the field of vision. However, Knight and Baber (2007) have questioned this location or at least raised concerns. Recording shoulder and upper arm muscle activity and measuring perceptions of exertion while participants interacted with arm mounted computers of different weights they found that the mere act of holding the arm in an appropriate posture to interact with an arm mounted computer was sufficient to exceed recommended levels of muscle activity for sustained activity . In addition it induced symptoms of fatigue after only two minutes where the addition of weight in the form of the mounted technology compounded this physical effect.

Figure 3. SenseWear device from BodyMedia[3]

Reducing Size and Separating Components

The discussion so far has suggested ways in which weight and placement of loads on the body can lead to problems The implication is that most wearables take the form factor of the ‘brick-on-the-back’ (as illustrated by Figure 2). It is possible to reduce the size of components further by removing the interaction and display components. Thus, the Bluetooth headset for a mobile phone effectively reduces the user interface (for talking on the phone) to a unit that clips on the ear. Various MP3 players can be worn on the upper arm while jogging (see comment about usability). Alternatively, there has been much interest in the use of ‘active badges’ that signal a person’s location to a network (Want et al., 1992). These devices can be made small enough to be worn as badges or incorporated into items of clothing, for example Schmidt et al. (1999) report a device, mounted in the wearer’s tie, which detects changes in ambient sound levels or wearer movement. In these examples, the ‘on-body’ part of the computer system is basically a sensor and/or transmitter and/or receiver that link the wearer to a network of other technologies. BodyMedia have developed a device that places sensors on the person’s upper arm in order to record data relating to everyday activity (in much the same way that an MP3 player is a data display device) and needs to be connected to a computer in order to process the tunes on it (see Figure 3). However, it makes sense to ask whether reducing the size of the devices will reduce or even eliminate these problems. Physical discomfort may arise from pressure of the device on the body. The pinnar is very sensitive and many users complain of discomfort around the ear from using ear plugs and ill-fitting head phones. HMDs place pressure around the forehead and crown. Even the small displays produced by Microoptical (see Figure 4) can cause discomfort around the ears and nose; especially given that the glasses they come with are hard plastic, not specifically fitted for the individual wearer. Items attached around the arm, specifically during physical activity (i.e., for jogging), have to be attached tightly so that they do not bounce against the arm or slip down, as such they may result in discomfort and result in vasoconstriction leading to sensations of numbness and tingling in the lower arm and hand, not to mention that the display is not easily viewable when the device is in position which raises usability issues.

Perceptual Impacts of Wearing Computers

A wearable computer can be considered a 'perceptual prosthesis' in the ways that it can provide additional information to, or enhance perceptual capability of, the wearer. Information provision can simply mean allowing the wearer access to information that is not currently present in the world, for example through a visual or audio display. The underlying concept in many of these applications is that the wearer's perception of the world can be augmented or enhanced (Feiner et al., 1998). Before considering this argument, we can consider the MP3 player as a device that can present media that is added to the world (in the form of a personal soundtrack to the person's activity). To some extent this could be considered as form of augmenting perception. What makes augmented reality different from simply displaying information is that (in most systems) the information is presented in accordance with the context in which the person is behaving. Such information could be in the form of an opaque visual display or could be visual information overlaid onto the world. In basic systems, the information can be called up by the wearer, or pushed from another source, for example a radio or telephone link. In more sophisticated systems, the information is presented on the basis of the computer's interpretation of 'context.'

While augmented reality displays could be beneficial, there are potential problems associated with the merging of one source of information (the computer display) with another (the world). Contemporary wearable computers tend to combine monocular head-mounted displays (see Figure 4) with some form of interaction device, for example keyboard, pointing device or speech. NRC (1997) point out that monocular head-mounted displays could suffer from problems of binocular rivalry, that is, information presented to one eye competes for attention with information presented to the other, which results in one information source becoming dominant and for vision to be directed to that source. A consequence of this phenomenon is that the wearer of a monocular display might find it difficult to share attention between information presented to the 'display eye' and information seen through the 'free eye.' This problem is compounded by the field of view of such displays.

Figure 4. Monocular head-mounted displays can restrict field of view

The field of view for the commercial monocular HMDs ranges from 16°-60°, which is considerably less than that of normal vision (around 170° for each eye, Marieb 1992). Narrow field of view can degrade performance on spatial tasks such as navigation, object manipulation, spatial awareness, and visual search tasks. Restrictions on field of view will tend to disrupt eye-head coordination and to affect perception of size and space (Alfano & Michel, 1990). One implication of a restricted field of view is that the wearer of a see-through HMD will need to engage in a significant amount of head movement in order to scan the environment (McKnight & McKnight, 1993). Seagull and Gopher (1997) showed longer time-on-task in a flight simulator when using a head down visual display unit than with a head-mounted, monocular display. Thus, it appears that a monocular display might impair performance. Apache helicopter pilots currently wear monocular, head-mounted displays and a review of 37 accidents concluded that 28 of these accidents could be attributed to wearing of the display (Rash et al., 1990).

Computer Response to Physical Activity

While it is possible that wearing technology affects user performance, it is also possible that the physical activity of the person can affect the computer, for example through the use of sensors to recognize actions and use this recognition to respond appropriately.

Given the range of sensors that can be attached to wearable computers, there has been much interest in using data from these sensors to define and recognize human activity. This has included work using accelerometers (Amft et al., 2005; Junker et al., 2004; Knight et al., 2007; Ling & Intille, 2004, Van Laerhoven & Gellersen, 2004; Westeyn et al., 2003) or tracking of the hand (Ogris et al., 2005). The approach is to collect data on defined movements to train the recognition systems, and then use these models to interpret user activity. At a much simpler level, it is possible to define thresholds to indicate particular postures, such as sitting or standing, and then to use the postures to manage information delivery (Bristow et al., 2004). The use of data from sensors to recognize human activity represents the merging of the research domains of wearable computers with that of pervasive computing, and implies the recognition not only of the actions a person is performing but also the objects with which they are interacting (Philipose, et al., 2004; Schwirtz & Baber, 2006).

An area of current interest for the wearable or ubiquitous computing communities is the interpretation of human movement related to maintenance and related activity. The artifacts with which the person is interacting could be instrumented. For example, a simple approach is to fit switches on components (in a self-assembly furniture pack) and for the user to depress the switches when each components is handled (Antifakos et al., 2002). A less intrusive approach would be to fit radio frequency identification person is using (Scwhirtz et al., 2006) and to use the activation of these tags to infer user activity. This requires that either the components or the tools be adapted to the task. The decreasing costs of RfiD suggest that, within a few years, tags will be universally used in a wide range of consumer products. Alternatively, the activity of the user could be taken to develop predictive models in order to infer the activity that is being performed. For some activities it might be sufficient to use data from very generic sensors, such as microphones, to collect data to define actions (Ward et al., 2006), while in other it might be necessary to rely on more specific sensors, such as the use of accelerometers to define movements that are characteristic of assembly and maintenance tasks (Schwirtz & Baber, 2006; Westyn et al., 2003). Steifmeier et al. (2006) show how tracking the motion of the hand (e.g., using ultrasonic tracking and inertial sensors) can be used to define specific types of movement that relate to maintenance tasks, such as spinning a wheel or rotating the pedals, unscrewing a cap or using a bicycle pump. In a related domain, there is some interesting work on the collection of activity data relating to nursing, for example through a combine infra-red proximity sensing and accelerometers (Noma et al., 2004).

USING WEARABLE COMPUTERS

Given that a wearable computer provides a means of 'anytime, anywhere' access to information, most forms of cognitive activity have been mooted as possible areas that can be supported. In this section, we consider three areas: (i.) supporting memory; (ii.) supporting navigation; and (iii.) information search and retrieval.

Supporting Memory

A common example used to illustrate the benefits of a wearable computer is what can be termed the 'context-aware memory.' Imagine you are attending a conference (or any large social gathering) and having to remember someone's name. Systems using some form of badge or face-recognition have been proposed to help with such situations; the computer would register the person and provide you with name and some additional details about the person, for example when you last met, what research interests are listed on their web-page, where they work, and so forth. There has been little research on whether and how these systems improve memory, and this example points to a possible confusion between supporting processes involved in recalling information from memory and the provision of contextually-relevant information. An associated question is whether wearable computers (particularly having a head-mounted display) positioned on the eye can have an impact on recall. The analogy is with the tourist watching a parade through the lens of a video camera—does the act of recording something weaken the ability to process and recall information? Baber et al. (2001) use a search task coupled with surprise recall to show that, in comparison with not using any technology, participants using a digital camera and wearable computer conditions showed lower performance, and that overall the wearable computer showed the biggest impairment in recall. There are many reasons why interruption at initial encoding can limit the ability to remember something, and the question is whether the head-mounted display serves to interrupt encoding; either due to distraction (with a host of information appearing on the screen), or through limitations of field of view, or for some other reasons.

Navigation and Way-Finding

Wayfinding requires people to travel through the world in order to reach specific locations. Thus, there is a need to manage both the act of traveling (for wearable computers this usually consists of walking) and relating a view of the world to the defined location. Sampson (1993) investigated the use of monocular, head-mounted displays for use when walking. Participants were presented with either 'spatial' or alphanumeric information and required to traverse paths with or without obstacles. In general, participants performed equally well when standing or when traversing paths without obstacle, but were significantly worse when obstacles were present. Thus, the need to maintain visual attention on display and environment can be seen to impair performance which suggests that the amount and type of information which can usefully be presented on such displays needs to be very limited. As global positioning systems (GPS) become smaller, cheaper, and more accurate, there has been an increase in their application to wearable computers. Seager and Stanton (2004) found faster performance with a paper map than GPS, that is routes completed faster. This was due, in part, with the increase in time spent looking at the GPS display and the number of updates made on the digital view (presumably because the GPS was moving a marker along the map to show the participants location). Participants also were less likely to orientate the digital view in the direction of travel (possibly because they might have assumed that the display would orient to their direction of travel rather than North up). Studies into the effect of perspective views, that is aligning a map with direction of travel, have not shown significant performance advantage over 2D views to date (Suomela et al., 2003), although this approach does seem beneficial in systems that support wayfinding in moving vehicles (Aretz, 1991). Systems that overlay routes onto the head-mounted display (Figure 5) could also assist in simple wayfinding tasks.

The use of visual support for navigation requires the user to divide attention between the environment and a visual display. In terms of the background or foreground of activity (discussed by Hinckley et al., 2005), this essentially places all the tasks in the users foreground. It is possible that navigation prompts could be provided using auditory cues. In simple terms, the heading could be indicated by varying the parameters of simple 'beeps,' for example changing in pitch or in intensity as the person deviates from a path (rather like auditory glide-slope indicators in aircraft). More recent developments have replaced the simple 'beeps' with music. In this way, a more subtle (background) form of cueing can be achieved to useful effect. The music could be manipulated to vary quality with deviation from a path, for example through distortion (Strachan et al., 2007, 2005), or through modifying the panning of music in stereo presentation (Warren et al., 2005).

Finding and Retrieving Information

Having the ability to access information as you need it is a core concept of wearable computer research. Early examples had the user enter queries and view information on a monocular display. This approach of having the user ask for information has been superseded by having the computer push information, on the basis of its interpretation of context. In such examples, the benefit of wearing the computer comes from its permanent presence and state of readiness, allowing access either of data stored on the computer or via the World Wide Web. Rhodes and Starner (1997) describe the 'remembrance agent,' which monitors the information that a user types into a computer and makes associations between this information and data it has stored. Obviously this creates an overhead on the user, in terms of the need to type information into the computer. However, it is only a relatively small modification to replace the typed entry with speech recognition (see Pham et al., 2005). While the 'agents' in either of these examples can run on desktop computers, it is the fact that they are continuously running on the computer worn by the user that makes them interesting. The role of the wearable computer in these examples is to discretely run searches in the background and alert the wearer to interesting links and associations between the current topic of conversation (or typing) and information to which the computer has access. It might also be useful for the computer to track user activity and then to either record patterns of activity (in order to refine its model of context) or to offer information relevant to the user. Thus, a very common application domain for context-aware, wearable computer research is the museum visitor (Sarini & Strapparava, 1998). The idea is that when the visitor stands at a certain location, say in front of a painting, the computer offers information relating

Figure 5. Displaying a route overlay on a head-mounted display

to the objects in that location. The manner in which the location (and other contextual factors) is used to select information and the manner in which the information is presented to the user might vary across applications. The examples considered thus far present the wearable computer as a device that is able to 'push' potentially relevant information to the wearer. However, rather than simply overloading the wearer with information, the role of the agents or context-awareness is to manage and tailor the information to the person.

Impact on User Performance

While there are many applications reported in the literature, there is a surprising lack of research into how effective these applications are in improving or otherwise changing user performance. One domain in which wearable computers have received both interest and support has been in maintenance (Mizell, 2003). A study exploring the use of head-mounted displays to support maintenance work show that performance could be improved, providing information was displayed in an appropriate format (Kancler et al., 1998). However, other studies have been more equivocal. Tasks requiring participants to follow instructions on a wearable computer or printed on paper have shown both the wearable computer (Baber et al., 1998) and the paper (Baber et al., 1999c; Ockerman et al., 1997; Siegel & Bauer, 1997) to lead to superior performance. One explanation of these differences lies in the design of the information that was presented, for example the manner in which information is presented can impact on overall performance times (Baber et al., 1999b; Sampson et al., 1993). One implication of these studies is that participants using the wearable computer tend to follow the same sequence of tests (as defined on the visual display), whereas the participants in the paper condition would order the tests as they saw fit (Baber et al., 1999b, Ockerman & Pritchett, 1998).

In terms of interacting with wearable computers and the appropriate devices to use, there has been very little work to date. While one might assume that the optimal interaction techniques would be ones that support hands-free interaction, such as speech recognition, studies suggest that walking has a negative impact on speech recognition performance (Oviatt, 2000; Price et al., 2004). In terms of entering data, Thomas et al. (1997) showed that a forearm mounted QWERTY keyboard led to superior performance over a five-button chording device or a virtual keyboard controlled using an isometric button. However, one might question the recommendation of a forearm mounted device, based on consideration of musculoskeletal strain. In terms of selecting objects on a display, Thomas et al. (1998) found that a touchpad mounted on the forearm was preferred by users, but that one mounted on the thigh lead to superior performance when sitting, kneeling or standing. Thus, the mounting of a pointing device can have a bearing on performance (although one might question whether pointing is an appropriate means of performing selection tasks on a wearable computer). Zucco et al. (2006) considered the performance of 'drag and drop' tasks while stationary and whilst walking, using different devices. They found that a gyroscopic mouse lead to best performance while stationary, but that touchpad or trackball were lead to better performance when the user was walking (and that all devices were superior to the Twiddler keypad).

Rather than seeing these studies as recommendations for specific interaction devices, I feel that they illustrate that the relationship between the activity that a person is performing and the demands of the ongoing activity in the environment interact in ways that require careful planning in the design of wearable computers. This brief discussion raises questions on one might develop standards for the deployment of wearable computers in these (and related) applications, and also what level of performance improvement one might expect from this technology.

SMART CLOTHING AND TEXTILES

It is worth mentioning the efforts to incorporate at least some aspects of technology into clothing and textiles (Berzowska, 2005). This could then take the form of clothing that has been adapted to incorporate some of the components (see Figure 6).

Alternatively, the textiles can exhibit some form of 'smartness.' Often this refers to "Textiles that are able to sense stimuli from the environment, to react to them and adapt to them..." (Van Langehoven & Hertleer, 2004). Taking these terms separately, we can ask how could textiles sense, react and adapt. In terms of sensing, there is a wide range of possible approaches, including thermal, chemical, mechanical, as well as biosensors. Buechley (2006) shows how simple off-the-shelf sensors and actuators can be incorporated into items of clothing. In particular, this work, following the earlier work of Post and Orth (1997) demonstrates how fabrics can be knitted or woven to provide some electrical conductivity, and then fitted with components. In contrast, other researchers report ways in which it is possible to use to electrical properties of prepared fabrics, such as change in resistance, inductance or capacitance (Wijesiriwardana et al., 2003, 2004), to incorporate specific sensing capabilities, such as stretching (Farringdon et al., 1999; Huang et al., 2006) or impact (Lind et al., 1997). What has yet to be fully realized from this smart textile work is that manner in which the user would interact with the 'computing' aspects in order to entry data or perceive displayed information. Thus, much of the research is still concerned with the development of the 'bus' onto which sensors, processors, displays, batteries, and so forth can be mounted.

EMOTIONAL IMPACT OF WEARING COMPUTERS

There are three broad categories of impact that will be considered in this section. The first concerns the emotional response to wearable computers by the people wearing these devices and the people with whom they interact. The second concerns the response of the computer to the emotions of the wearer, that is affective computing. The third concerns the manner in which collaboration can be supported within a group of wearable computer wearers.

While head-mounted displays have been used for many years in the domain of military aviation, they have yet to find widespread use on the ground. Most cinema goers will be familiar with concepts of wearable computers from science fiction films and might feel uncomfortable with 'cyborgs' in their midst. This discomfort could be particularly acute in stressful situations (as one of the paramedics in an earlier study pointed out to us, *If you're just coming round from a heart attack, the last thing you'd expect to see if some robot headed bloke trying to take your pulse*). As there are so few commercial applications of wearable computers, these devices still represent something of a novelty and there has been very little research into how people might respond to people wearing such devices. One collection of anecdotal evidence can be found in Thad Starner's Ph.D. Starner completed his Ph.D. at MIT, and with Brad Rhodes and Steve Mann developed a variety of wearable computers that they wore

Figure 6. Incorporating consumer electronics into clothing—the Philips-Levi Strauss jacket[4]

for prolonged periods of time. This meant that they would encounter members of the public on a regular basis and Starner recorded some of the reactions. A common response was to assume that the wearable computer was simply some form of familiar technology, perhaps a very sophisticated video camera or an advanced form of video game console or a medical device. The implication of this is that people might try to explain the technology using a familiar mental model but, as Starner points out, people have yet developed a mental model of a wearable computer on the basis of familiarity and exposure. While this is partly a matter of the unfamiliar appearance of the wearer it is also related to the manner in which the wearer interacts with the device. For example, Starner points out that when you ask someone the time they will raise their wrist in order to consult their watch, but with a head-mounted display one might simply glance up at the screen. This means that the physical behavioral cues might be far less obvious. This is, of course, similar to the way in which Bluetooth headsets allow people to speak on their mobile telephone in a manner that makes it looks as if they are speaking to themselves (indeed, some users of Bluetooth headsets make gestures that look as if they are holding a phone, for example holding the microphone or holding their hands near their faces). There will be a trade-off between the familiarity of wearable computers, their appearance, and the ability of people to explain their use. However, there is an additional factor at play here (which is also hinted at by the Bluetooth headset discussion), and that is that the wearable computer provides information to an individual in a highly individualized manner—it is not possible for other people to see what the wearer is looking at or guess what they are doing. An analogy can be drawn in this instance with the person on a train reading a humorous book at laughing out loud—the other people feel uncomfortable because they can not read the what the person is reading. The removal of the observer from the source of information can be disconcerting and can lead to such comments as (from Starner) “we can’t tell if you’re talking about us behind our backs” or “when you wear your display, how can I tell if you are paying attention to me or reading your e-mail?”

One implication of the physical appearance and individualized interaction of wearable computers is the sense that people who wear such technology are different from ‘normal’ people. As Sheridan et al. (2000) note, people wearing computers could be perceived (by themselves or by the people around them) as a ‘cyborg community’ that is different from other people. With the use of communication and networking capabilities, it is possible for a group of wearable computer users to be able to share information and maintain contact as a community (Wellman, 2001). However, we should be careful to distinguish between the ability to maintain contact with other people (which one can do easily with mobile telephones, even to the extent of setting up talk groups) and the ability to share the wide range of information that wearable computing can support. This could simply mean the sharing of the same documents or video, but could also allow new forms of collaboration, sharing and exchange of information.

Context could be defined by changes in the physiological state of the wearer (Picard, 1997). This requires a more intimate means of recording data from the wearer, perhaps through monitoring of pulse or heart activity. The use of ‘context’ to initiate image capture has been demonstrated by several projects; most notably in Healey and Picard’s (1998) ‘StartleCam,’ in which changes in galvanic skin response (GSR) was used to trigger image capture. There has been surprisingly little attempt at extending this work in the years since it was reported, although over uses of ‘context’ in image capture have explored the use of ambient sound which is captured at the same time as the image (e.g., Frolich & Tallyn, 1999; Ljunblad et al., 2004). Bristow et al. (2005) used a set of context identifiers to take still images when ‘context’ changed, and showed that these were surprisingly consistent with photographs taken by humans. In addition to the computer ‘sensing’ the physiological responses of the wearer (and hence drawing some inference as to the affective state), it is also possible to infer the state of people with whom the wearer is interacting in order to develop ‘emotionally intelligent interfaces.’ By monitoring the changing facial expressions of ones conversational partner, it might be possible to provide support for people who find it difficult to judge the emotional response, for example people with Autism (El Kaliouby & Robinson, 2003).

DISCUSSION

Wearable computers continue to raise many significant challenges for human factors research. These challenges involve not only cognitive aspects of presenting information but also perceptual aspects of displaying the information against the backdrop of the everyday environment and physical aspects of mounting the devices on the person. This chapter has overviewed some of the developments in the field and offered some consideration of how these human factors can be considered. While the field is largely motivated by technological advances there is a need to carefully ground the developments in the physical and cognitive characteristics of the humans who are intended to wear them.

REFERENCES

Alfano, P.L., & Michel, G.F.(1990). Restricting the field of view: Perceptual and performance effects. *Perceptual and Motor Skills, 70,* 35-45.

Amft, O., Junker, H., & Troster, G. (2005). Detection of eating and drinking arm gestures using inertial body-worn sensors, *Ninth International Symposium on Wearable Computers ISWC 2005* (pp.160-163). Los Alamitos, CA: IEEE Computer Society.

Antifakos, S., Michahellis, F., & Schiele, B. (2002). Proactive instructions for furniture assembly. *UbiComp 2002 4th International Conference on Ubiquitous Computing* (pp. 351-360). Berlin: Springer.

Aretz, A.J. (1991). The design of electronic map displays. *Human Factors, 33*(1), 85-101.

Baber, C. (1997). *Beyond the desktop: designing and using interaction devices.* San Diego, CA: Academic Press.

Baber, C., Haniff, D.J., Knight, J., Cooper, L., & Mellor, B.A. (1998). Preliminary investigations into the use of wearable computers. In R. Winder (Ed.), *People and computers XIII* (pp. 313-326). London: Springer-Verlag.

Baber, C., Haniff, D.J., & Woolley, S.I. (1999a). Contrasting paradigms for the evelopment of wearable computers. *IBM Systems Journal, 38*(4), 551-565.

Baber, C., Knight, J., Haniff, D., & Cooper, L. (1999b). Ergonomics of wearable Computers. *Mobile Networks and Applications, 4,* 15-21.

Baber, C., Arvanitis, T.N., Haniff, D.J., & Buckley, R. (1999c). A wearable computer for paramedics: studies in model-based, user-centred and industrial design, In M.A. Sasse and C. Johnson (Eds.), *Interact'99* (pp. 126-132). Amsterdam: IOS Press.

Baber, C., Sutherland, A., Cross, J., & Woolley, S. (2001). A wearable surveillance system: Implications for human memory and performance. In D. Harris (Ed.), *Engineering psychology and cognitive ergonomics industrial ergonomics, HCI, and applied cognitive psychology, 6,* (pp. 285-292). Aldershot: Ashgate.

Bainbridge, L. (1987) Ironies of automation. In J. Rasmussen, J. Duncan & J. Leplat (Eds.), *New technology and human errors* (pp. 271-286). New York: Wiley

Bass, L. (1996). Is there a wearable computer in your future? In L.J. Bass and C. Unger (Eds.), *Human computer interaction* (pp. 1-16). London: Chapman and Hall.

Bass, L., Siewiorek, D., Smailagic, A,. & Stivoric, J. (1995). On site wearable computer system. *CHI'95* (pp. 83-84). New York: ACM.

Berzowska, J. (2005). Electronic textiles: wearable computers, reactive fashions and soft computation. *Textile, 3,* 2-19.

Bodine, K., & Gemperle, F. (2003). Effects of functionality on perceived comfort of wearables, *Digest of Papers of the 7th International Symposium on Wearable Computers,* (pp. 57-61). Los Alamitos, CA: IEEE Computer Society.

Buechley, L. (2006) A construction kit for electronic textiles. In *Digest of Papers of the 10th International Symposium on Wearable Computers* (pp. 83-90). Los Alamitos, CA: IEEE Computer Society.

Bristow, H., Baber, C., Cross, J., Knight, J., & Woolley, S. (2004). Defining and evaluating context for wearable computers *International Journal of Human Computer Studies*, 60, 798-819.

Clark, A. (2006). Natural born cyborgs? *Edge. Retrieved from* www.edge.org/3rd_culture/clark/clark_index.html

El Kaliouby, R., & Robinson, P. (2003). The emotional hearing aid—An assistive tool for autism, *HCI International*

Farringdon, J., Moore, A.J., Tilbury, N., Church, J., & Biemond, P.D. (1999). Wearable sensor badge and sensor jacket for contextual awareness. *Digest of Papers of the 3rd International Symposium on Wearable Computers* (pp. 107-113). Los Alamitos, CA: IEEE Computer Society.

Feiner, S., MacIntyre, B., Hollerer, T., & Webster, A. (1997). A touring machine: Prototyping 3D mobile augmented reality systems for exploring the urban environment. *Digest of Papers of the 1st International Symposium on Wearable Computers*, (pp. 74-81). Los Alamitos, CA: IEEE Computer Society.

Frolich, D., & Tallyn, E. (1999). AudioPhotography: Practice and prospects. *CHI'99 Extended Abstracts* (pp. 296-297). New York: ACM.

Gemperle, F., Kasabach, C., Stivoric, J., Bauer, M., & Martin, R. (1998). Design for Wearability. *Digest of Papers of the 2nd International Symposium on Wearable Computers* (pp. 116-123). Los Alamitos, CA: IEEE Computer Society.

Healey, J., & Picard, R.W. (1998). StartleCam: a cybernetic wearable computer, *Proceedings of the 2nd International Symposium on Wearable Computers* (pp. 42-49). LosAlamitos, CA: IEEE Computer Society.

Hinckley, K., Pierce, J., Horvitz, E., & Sinclair, M. (2005). Foreground and background interaction with sensor-enhanced mobile devices. *ACM Trans. Computer-Human Interaction, 12*(1), 31-52.

Huang, C-T., Tang, C-F., & Shen, C-L. (2006). A wearable textile for monitoring respiration, using a yarn-based sensor. *Digest of Papers of the 10th International Symposium on Wearable Computers*, Los Alamitos, CA: IEEE Computer Society, 141-142

Junker, H., Lukowicz, P., & Troster, G. (2004). Continuous recognition of arm activities with body-worn inertial sensors. *Eighth International Symposium on Wearable Computers ISWC'04* (pp.188-189). Los Alamitos, CA: IEEE Computer Society.

Kancler, D.E., Quill, L.L., Revels, A.R., Webb, R.R., & Masquelier, B.L. (1998). Reducing cannon plug connector pin selection time and errors through enhanced data presentation methods. *Proceedings of the Human Factors and Ergonomics Society 42nd Annual Meeting* (pp. 1283-1290). Santa Monica, CA: Human Factors and Ergonomics Society.

Knight, J. F., Baber, C., Schwirtz, A., & Bristow, H. W. (2002). The comfort assessment of wearable computers. *The Sixth International Symposium of Wearable Computers* (pp. 65-72). Los Alamitos, CA: IEEE Computer Society.

Knight J.F., & Baber, C. (2004). Neck muscle activity and perceived pain and discomfort due to variations of head load and posture. *Aviation, Space and Environmental Medicine, 75*, 123-131.

Knight, J. F., & Baber, C. (2005). A tool to assess the comfort of wearable Computers. *Human Factors, 47*(1), 77-91.

Knight, J. F., & Baber, C. (2007). Assessing the physical loading of wearable computers. *Applied Ergonomics, 38*, 237-247.

Knight, J.F., Williams, D.D., Arvanitis, T.N., Baber, C., Sotiriou, S., Anastopoulou, S. et al. (2006). Assessing the wearability of wearable computers. *Digest of Papers of the 10th International*

Symposium on Wearable Computers (pp. 75-82). Los Alamitos, CA: IEEE Computer Society.

Knight, J., Bristow, H.W., Anastopoulou, S., Baber, C., Schwirtz, A., &Arvanitis, T.N. (2007). Uses of accelerometer data collected from a wearable system. *Personal and Ubiquitous Computing 11*(2), 117-132.

Legg, S. J. (1985). Comparison of different methods of load carriage. *Ergonomics*, *28*(1), 197-212.

Lin, R., & Kreifeldt, J. G. (2001) Ergonomics in wearable computer design. *International Journal of Industrial Ergonomics*, *27*(4), 259-269.

Lind, E.J., Jayaraman, S., Park, S., Rajamanickam, R., Eisler, R., Burghart, G., et al. (1997). A sensate liner for personnel monitoring applications. *Digest of Papers of the 1st International Symposium on Wearable Computers* (pp. 98-107). Los Alamitos, CA: IEEE Computer Society.

Ling, B., & Intille, S. S. (2004). Activity recognition from user-annotated acceleration data. *Pervasive 2004* (pp. 1–17). Berlin: Springer.

Ljungblad, S., Hakansson, M., Gaye, L & Holmquist, L.E. (2004). Context photography: Modifying the digital camera into a new creative tool, *CHI'04 Short Papers*, (pp. 1191-1194). New York: ACM.

Mann, S. (1997). An historical account of 'WearComp' and 'WearCam' inventions developed for applications in 'Personal Imaging.' In *The First International Symposium on Wearable Computers* (pp. 66-73). Los Alamitos, CA: IEEE Computer Society.

Marieb, E. N. (1992). *Human anatomy and physiology* [2nd edition]. Redwood City, CA: Benjamin/Cummings.

McKnight, A.J., & McKnight, A.S. (1993). The effects of motor cycle helmets upon seeing and hearing. In *Proceedings of the 37th Annual conferences of the Association for the Advancement of Automotive Medicine* (pp. 87-89). Des Plaines, Ill.

Mizell, D. (2001). Boeing's wire bundle assembly project. In W. Barfield & T. Caudell (Eds.), *Fundamentals of wearable computers and augmented reality* (pp. 447-469). Mahwah, NJ: LEA.

Naya, F., Ohmura, R., Takayanagi, F., Noma, H., & Kogure, K. (2006). Workers' routine activity recognition using body movement and location information. *Digest of Papers of the 10th International Symposium on Wearable Computers* (pp. 105-108). Los Alamitos, CA: IEEE Computer Society.

Nigg, B.M., & Herzog, W. (1994). *Biomechanics of the musculoskeletal system*. England: Wiley

NRC. (1997). *Tactical display for soldiers*. Washington, DC: National Academy.

Noma, H., Ohmura, A., Kuwhara, N., & Kogure, K. (2004). Wearable sensors for auto-event recording on medical nursing—User study of ergonomic requirements. *Digest of Papers of the 8th International Symposium on Wearable Computers* (pp. 8-15). Los Alamitos, CA: IEEE Computer Society.

Ockerman, J.J., & Pritchett, A.R. (1998). Preliminary investigation of wearable computers for task guidance in aircraft inspection. *Digest of Papers of the 2nd International Symposium on Wearable Computers* (pp. 33-40). Los Alamitos, CA: IEEE Computer Society.

Ogris, G., Stiefmeier, T., Junker, H., Lukowics, P., & Troster, G. (2005). Using ultrasonic hand tracking to augment motion analysis based recognition of manipulative gestures. *ISWC 2005* (pp. 152-159). Los Alamitos, CA: IEEE Computer Society.

Oviatt, S. (2000). Taming recognition errors with a multimodal interface. *Communications of the ACM, 43*, 45-51.

Patterson, D. (2005). Fine grained activity recognition by aggregating abstract object usage. *Digest of Papers of the 9th International Symposium on Wearable Computers* (pp. 44-51). Los Alamitos, CA: IEEE Computer Society.

Pham, N.V., Terada, T., Tsukamato, M., & Nishio, S. (2005). An information retrieval system for supporting casual conversation in wearable computing environments. *Proceedings of the Fifth International Workshop on Smart Appliances and Wearable Computing* (pp. 477 – 483). Washington, DC: IEEE Computer Society.

Philipose, M., Fishkin, K. P., Perkowitz, M., Patterson, D. J., Fox, D., Kautz, H., & Hahnel, D. (2004). Inferring activities from interactions with objects. *Pervasive Computing 2004,* 50–57.

Picard, R. (1997). *Affective computing.* Cambridge, MA: MIT Press.

Post, E.R. & Orth, M. (1997). Smart fabric, or "wearable clothing." *Digest of Papers of the 1st International Symposium on Wearable Computers* (pp. 167-168). Los Alamitos, CA: IEEE Computer Society.

Price K. J., Min L., Jinjuan, F., Goldman, R., Sears, A., & Jacko, J.A. (2004). Data entry on the move: An examination of nomadic speech-based text entry. *User-centered interaction paradigms for universal access in the information society* (pp. 460-471). Berlin: Springer.

Rash, C.E., Verona, R.W., & Crowley, J.S. (1990). Human factors and safety considerations of night vision systems flight using thermal imaging systems. *Proceedings of SPIE–The International Society for Optical Engineering, 1290,* 142-164.

Rhodes, B.J., & Starner, T. (1997). Remembrance agent: A continuously running automated information retrieval system. *The Proceedings of The First International Conference on The Practical Application Of Intelligent Agents and Multi Agent Technology* (PAAM '96), pp. 487-495.

Rohaly, A. M., & Karsh, R. (1999). Helmet-mounted displays. In J. M. Noyes & M. Cook (Eds.), *Interface technology: The leading edge* (pp. 267-280). Baldock: Research Studies Press Ltd.

Sampson, J. B. (1993). Cognitive performance of individuals using head-mounted displays while walking. In *Proceedings of the Human Factors and Ergonomics Society 37th Annual Meeting* (pp. 338-342). Santa Monica, CA: Human Factors and Ergonomics Society.

Sarini, M., & Strapparava, C. (1998, June 20-24). *Building a user model for a museum exploration and information-providing adaptive system*. Proceedings of the Workshop on Adaptive Hypertext and Hypermedia. Pittsburgh, PA.

Sawhney, N., & Schmandt, C. (1998). Speaking and listening on the run: Design for wearable audio computing. *Digest of Papers of the 2nd International Symposium on Wearable Computers* (pp. 108-115). Los Alamitos, CA: IEEE Computer Society.

Schmidt, A., Gellerson, H-W., & Beigl, M. (1999). A wearable context-awareness component: Finally a good reason to wear a tie. *Digest of Papers of the 3rd International Symposium on Wearable Computers* (pp. 176-177). Los Alamitos, CA: IEEComputer Society.

Schwirtz, A., & Baber, C. (2006). Smart tools for smart maintenace. *IEE and MoD HFI-DTC Symposium on People in Systems* (pp. 145-153). London: IEE.

Seager, W., & Stanton, D. (2004, April). *Navigation in the city.* Ubiconf *2004,* Gresham College, London.

Seagull, F. J., & Gopher, D. (1997). Training head movement in visual scanning: An embedded approach to the development of piloting skills with helmet-mounted displays. *Journal of Experimental Psychology: Applied*, *3*, 163-180

Siegel, J., & Bauer, M. (1997). On site maintenance using a wearable computer system, *CHI'97* (pp. 119-120). New York: ACM.

Sheridan, J. G., Lafond-Favieres, V., & Newstetter, W. C. (2000). Spectators at a Geek Show: An Ethnographic Inquiry into Wearable Computing, *Digest of Papers of the 4th International Symposium on Wearable Computers* (pp. 195-196). Los Alamitos, CA: IEEE Computer Society.

Starner, T. (1999). *Wearable computing and contextual awareness.* Unpublished Ph.D. thesis, Masachusetts Institute of Technology, USA.

Stiefmeier, T., Ogris, G., Junker, H., Lukowicz, P. & Tröster, G. (2006). Combining motion sensors and ultrasonic hands tracking for continuous activity recognition in a maintenance scenario. *Digest of Papers of the 10th International Symposium on Wearable Computers* (pp. 97-104). Los Alamitos, CA: IEEE Computer Society.

Stein, R., Ferrero, S., Hetfield, M., Quinn, A., & Krichever, M. (1998). Development of a commercially successful wearable data collection system. *Digest of Papers of the 2nd International Symposium on Wearable Computers* (pp. 18-24). Los Alamitos, CA: IEEE Computer Society.

Strachan, S., Williamson, J., & Murray-Smith, R. (2007). Show me the way to Monte Carlo: density-based trajectory navigation. In *Proceedings of ACM SIG CHI Conference*, San Jose.

Strachan, S., Eslambolchilar, P., Murray-Smith, R., Hughes, S., & O'Modhrain, S. (2005). GpsTunes—Controlling navigation via audio feedback. *Mobile HCI 2005.*

Suomela, R., Roimela, K., & Lehikoinen, J. (2003), The evolution of perspective view in WalkMap. *Personal and Ubiquitous Computing, 7,* 249-262.

Tan, H. Z., & Pentland, A. (1997). Tactual displays for wearable computing. *Digest of Papers of the 1st International Symposium on Wearable Computers* (pp. 84-89). Los Alamitos, CA: IEEE Computer Society.

Thomas, B., Tyerman, S., & Grimmer, K. (1997). Evaluation of three input mechanisms for wearable computers. In *Proceedings of the 1st International Symposium on Wearable Computers* (pp. 2-9). Los Alamitos, CA: IEEE Computer Society.

Thomas, B., Grimmer, K., Makovec, D., Zucco, J., & Gunther, B. (1998). Determination of placement of a body-attached mouse as a pointing input device for wearable computers. *Digest of Papers of the 3rd International Symposium on Wearable Computers* (pp. 193-194). Los Alamitos, CA: IEEE Computer Society.

Van Laerhoven, K., &Gellersen, H. W. (2004). .Spine versus porcupine: A Study in distributed wearable activity recognition. In *Proceedings of the 8th IEEE International Symposium on Wearable Computers (ISWC'04),* (pp. 142-149). Los Alamitos, CA: IEEE Computer Society.

Van Langehoven, L., & Hertleer, C. (2004). Smart clothing: a new life. *International Journal of Clothing Science and Technology, 13,* 63-72.

Want, R., Hopper, A., Falcao, V., & Gibbons, J. (1992). The active badge location system. *ACM Transactions on Information Systems, 10*(1), 91-102.

Ward, J.A., Lukowicz, P., & Tröster, G. (2006). Activity recognition of assembly tasks using using body-worn microphones and accelerometers. *IEEE Transaction on Pattern Analysis and Machine Intelligence.*

Warren, N., Jones, M., Jones S., & Bainbridge, D. (2005, April 2-7). *Navigation via continuously adapted music*. Extended Abstracts ACM CHI 2005, Portland, Oregon.

Wellman, B. (2001). Physical place and cyberspace: The rise of personalized networks. *International Journal of Urban and Regional Research, 25,* 227-252.

Westeyn, T., Brashear, H., Atrash, A. & Starner, T. (2003). Georgia Tech Gesture Toolkit: Supporting Experiments in Gesture Recognition, *ICMI'03*, ??

Wijesiriwardana, R., Dias, T. & Mukhopadhyay, S. (2003). Resisitivefiber meshed transducers. *Digest of Papers of the 7th International Symposium on Wearable Computers* (pp. 200-205). Los Alamitos, CA: IEEE Computer Society.

Wijesiriwardana, Mitcham, K. R., & Dias, T. (2004). Fibre-meshed transducers based on real-time wearable physiological information monitoring system. *Digest of Papers of the 8th International Symposium on Wearable Computers* (pp. 40-47). Los Alamitos, CA: IEEE Computer Society.

Zingale, C., Ahlstrom, V., & Kudrick, B. (2005). *Human factors guidance for the use of handheld, portable and wearable computing devices*. Atlantic City, NJ: Federal Aviation Administration, reprt DOT/FAA/CT-05/15

Zucco, J. E., Thomas, B. H. & Grimmer, K. (2006). Evaluation of four wearable computer pointing devices for drag and drop tasks when stationary and walking. *Digest of Papers of the 10th International Symposium on Wearable Computers* (pp. 29-36). Los Alamitos, CA: IEEE Computer Society.

KEY TERMS

Activity Models: Predictive models of human activity, based on sensor data

Augmentation Means: Devices that can augment human bevaior—a term coined by Doug Engelbart, and covering: *Tools & Artifacts:* the technologies that we use to work on the world which supplement, complement or extend our physical or cognitive abilities; *Praxis:* the accumulation and exploitation of skills relating to purposeful behavior in both work and everyday activity*; Language:* the manipulation and communication of concepts; *Adaptation:* the manner in which people could (or should) adapt their physical and cognitive activity to accommodate the demands of technology.

Comfort: Subjective response to wearing a wearable computer (ranging from physical loading to embarrassment)

Context-Awareness: The capability of a device to respond appropriately to changes in a person's activity, environment, and so forth.

Form-Factor: The overall size (and shape) of a device

Sensors: Devices that produce digital output in response to some change in a measured parameter, for example dependent on environmental change or on user activity

Wearable Computers: Devices worn on the person that provided personalized, context-relevant information

ENDNOTES

1 http://www.intel.com/research/exploratory/motes.htm

2 Bainbridge (1987) argued that full automation can lead to the ironic situation that, the role of the human operator is to intervene when something goes wrong. However, the automation is such that the human is locked out of the process and has little understanding as to what is happening. Consequently, the human will not be able to intervene in an informed and efficient manner. Ultimately, it means that, by designing the human out of the system, the potential for a flexible and intelligent response to unknown situations is lost.

3 http://www.bodymedia.com/main.jsp

4 http://www.extra.research.philips.com/pressmedia/pictures/wearelec.html

Chapter XI
The Garment as Interface

Sabine Seymour
Moondial Fashionable Technology, Austria

ABSTRACT

This chapter focuses on the surface of a smart garment as a dynamic interface. The use of the garment's surface as an interactive display opens up an array of new applications. Novel developments in interactive and wearable textile surfaces for garments display data from input sources like sensors and cell phones. The integration of these surfaces into the garments is evaluated regarding wearability and the wearer's interaction. This chapter provides a list of considerations for human interaction with smart garments and dynamic visual interfaces, which are an essential tool to design usable, smart garments.

ELECTRONIC TEXTILES

Electronic textiles or wearables result from the integration of technology into a textile, a garment, or a wearable object. Such objects or devices can either be embedded into the skin or into a textile or wearable material, or be portable. "An electronic textiles—or smart fabric/textile—refers to a textile substrate that incorporates capabilities for sensing (biometric or external), communication (usually wireless), power transmission, and interconnection technology to connect sensors and microprocessors to be networked together within the fabric" (Berzowska, 2005, p. 60). Wearable technologies are closely related to electronic textiles. The term wearable technologies covers in particular electrical engineering, physical computing, and wireless technologies. Electronic textiles and wearable technologies are literally interwoven. Intelligent garments have an enhanced functionality through embedded technologies. Integrated sensors monitor vital signs, built-in speech-recognition systems allow for an interface independent from a physical interaction, and embedded wireless systems enable hands-free communication. Intelligent garments is functional clothing constructed with textiles and materials that are considered smart. "Fibre sensors, which are capable of measuring temperature, strain/stress, gas, biological species and small, are typical smart fibres that can be directly applied to textiles" (Tao, 2001, p. 4). Smart textiles are capable of reacting to a stimulus with or without the use of competition. Outlast explains that PCMs—phase changing materials—can absorb, store, and release heat while the material changes from solid to liquid and back to solid.

H2: Dynamic Garment Interface

What is a dynamic interface of a garment? Technologies enrich the cognitive characteristics of our second skin—the surface of our garments. Currently, the surface of a garment is mostly used for static displays of information or for safety features. An example is the glow-in-the-dark function of running and biking gear to warn drivers on the road. The dynamic information, the output, can either be real-time or static. It can be controlled through a microprocessor or a simple non-computational input like light. The dynamic character of the surface, its colors, animation, lengths of appearance, subject, speed of movement, and so forth is influenced by the input. Though, the interface of the surface is not limited to a visual output. Its effectiveness can be achieved through an array of outputs for example motors or speakers. The focus of this paper is on the surface of the garment as a visual output devices, the garment as interface. The inputs can be many, and the interactions may vary, though the output is pre-defined through the surface. The objective is to describe the considerations for the design of an intelligent garment regarding the human interaction with the embedded electronic components, their placement, and the breath of functions that need to be considered in this novel field. It furthermore describes the importance of aesthetics in the design of a very personal yet public surface of the garment fueled with numerous preconceptions. Today's interface design in a ubiquitous environment is not restricted to two-dimensional displays stationary of computer monitors with the use of a mouse or mobile devices.

Multidisciplinary Character

The design of an intelligent garment is complex because of the breadth of disciplines needed for the development and because of the constraints the embedded technologies cause. A common vocabulary needs to be developed to allow for the many disciplines—like physical computing, fashion design, industrial design, wireless networking, software engineering, graphic design—to collaborate efficiently and fruitfully. Often the hands-on expertise of the craft—in particular garment construction—is not considered in the design process. A seamstress that understands the flow of electricity in a garment is rare and could soon be a know-how in high demand. "All too often projects covering this area fail fashion design—a flaw that often follows when engineers are dealing with the integration of technology in fashion. Conversely, where fashion designers who have no background in physical computing or programming work in the field, the actual technical integration is often flawed or absent" (Seymour, 2004, p. 13). A textile designer and an electrical engineer need to find a suitable common vocabulary in this novel field, as do all trades and disciplines involved.

The term wearable technologies covers in particular electrical engineering, physical computing, and wireless technologies.

EMBEDDED TECHNOLOGIES

Embedded technologies influence the wearability and comfort, the interaction system, and the aesthetic of the intelligent garment. If the functionality requires an active input by the wearer, the simplicity in understanding the technical features is key. Thus, the inclusion of wearable technologies know-how in the beginning stages of the design phase of a functional garment is central for its success. McCann, Hurford, and Martin (2005) describe the critical path during the design process of smart garments with end-user requirements on one hand and appropriate technology to fulfill such needs on the other hand. They identify the following items or processes: fiber/yarn, fabric, dying and finishing, coatings and laminates, body measurements/sizing and fitting, garment development, pattern development and grading, modeling, simulation and initial prototyping, integration of smart and wearable technologies, fitting, manufacture (cutting/bonding): mass and custom, distribution, display and point of sale: in-store and online, and end of life/recycling.

Inputs

Inputs, often sensors, answer where, who, what, when, how, and why. The huge variety of sensors makes it obvious how many possibilities there are in using them in the construction of intelligent clothing. For example, the inputs influence the user interaction extremely. Pressure is used for buttons and switches. The resistance varies with the application of force when using force-sensitive resistors. Pressure can also be produced through displacement. The resistance of bend sensors varies with the materials used. The firmer the material the harder it is to bend and as a result the interaction with the input varies. Other typical inputs are sensors such as temperature, proximity, magnetic hall sensors, and humidity sensors and do not require an active wearer interaction The environment delivers a number of inputs like macro-particles in smoke, smell, or optical scattering. The use of the environment as an input is particularly popular when referring to pollution. The wearer offers a dynamic display as the output—a human billboard—for current environmental data and is intertwined with the environment. This presents an interesting aspect of interaction. It seems to be passive but the act of choosing to display the information already requires the wearer to interact with the garment and to make conscious, active decision. As a result the wearer is consciously and actively involved. Sensors for an active interaction are for example acoustic, photodiodes, optical, accelerometers, touch, capacitive, compass, and orientation sensors. Many inputs can be both active and passive depending on the way they are used. The translation from input to output depends on the envisioned functionality.

The Visual (Dynamic) Surface as Output

A garment is seen, felt, heard, and touched. It stimulates the five senses. This chapter primarily focuses on the output of data on the surface of the garment, which is the display of information on a visual interface. All other outputs, for example motors, buzzers, speakers, fog or smoke will not be addressed. The use of the garment's surface as an interactive display is researched at various institutions and subject of interest for many artists. In 2003 France Telecom developed prototypes of flexible color screens using light-emitting diodes (LEDs) integrated in clothing calling it an optical fiber flexible display (OFFD). The prototype allows downloading, creating or exchanging visual data via the appropriate Internet gateway. (Konar, Deflin, & Weill, 2005). Barbara Layne, a researcher at Hexagram in Montreal works on a project called 'Animated Textiles,' which integrates light-emitting diodes (LEDs) and electronic circuitry into the structure of hand woven fabrics. 'Electronic Plaid,' by International Fashion Machine, layers thermochromic pigments using textile-printing technologies on top of an electronic textile with resistors woven into patterns. Laura and Lawrence MacCary, an artist and retired engineer based in Seattle create conductive fibers by weaving lead into 36 amplifier circuits, which illuminate LEDs depending on where you touch the fabric. The art-

Table 1. Classification of inputs

Origin	**Input formats (examples)**
Person	pressure, motion, vital signs, speech
Environment/Ambience	light, humidity, sound, temperature, smoke

work is entitled ‘Dialectric: Connections.’ In 2006 Philips launched Lumalive, a dynamic wearable textile surface. Lumalive fabrics feature arrays of colored light-emitting diodes (LEDs) fully integrated into the fabric that make it possible to create garments that can show dynamic messages, graphics or multicolored surfaces.

There are many factors that influence the dynamic textile-based display on the surface of the smart garment. For example the time the thermochromic needs to react to heat generated by conductive fibers woven into the material or the selection of the color of LEDs that are embedded into the fabric.

1. The variables captured from input sources (for example from sensors) are software-based data and consequently allow computation.
2. The data that goes through the microprocessor, the brain of the garment, and the computation programmed obviously determine the output.

The research into flexible organic light emitting diodes (FOLEDs) for garments seems hampered due to new developments in textile-based flexible displays. However, a visual dynamic textile output does not have to be dynamic. Very simple visual outputs, like one LED, are often sufficient to inform the wearer. Humans are already conditioned through their every day lives. Red means stop, green means go. Therefore a simple visual key through color might often be enough. If the heart rate of a heart patient is exceeding a certain beat an integrated red LED lights up indicating that the wearer of the intelligent garment needs to rest. This indication can be hidden in an area primarily visible for the wearer or obviously placed for care personnel to be able to monitor more easily individual patients without restraining them to a machine with a monitor.

Microprocessors

The microprocessor is necessary for the computation of the data derived from the input sources. Through computation the outputs are addressed on either the garment itself—referring to a BAN (body area network)—or the data is transferred wirelessly to a larger computer that processes the data and either gets back to the garment or keeps the data depending on the purpose of the application. Microprocessors or microcontrollers are the brains of the intelligent garment or wearable object. The single-chip computer can run and store a program but despite major research has still limited possibilities. Today the Arduino prototyping board seems to be the microprocessor of choice at universities and research settings where a quick and simple prototype is necessary. Such prototyping is very important to understand the functionalities and the interaction necessary by the user. Failure of an intelligent garment can thus be avoided. Microprocessors are becoming more and more relevant in the use for everyday garments due to their size, flexibility, and energy consumption, as are chips in everyday household appliances.

Issues Around Networks

The regional differences in wireless communication networks even just five years ago made it difficult to develop for a global audience. Today, however, the global collaborations in developing common standards make it easier to conceptualize and design embedded systems that endure geographical changes and standards. The need for flexibility leads to an obvious step in designing modular systems than can be removed and interchanged easily, thus influencing the architecture of the garment. The user needs to be able to switch off the network. An empowerment of the user is necessary for the success of such systems. The user might not be the actual wearer in particular when dealing with elderly care or neurologically sick patients like Alzheimer. However, the wearer needs to be in charge, not the machine.

Typical connectors in fashion design like buttons, zippers, grommets, snaps, or hooks and eye can easily be transformed into connectors and conductors of electricity, for example metal-based snaps. Other conductive connectors are conductive threads and fabric, conductive Velcro, conductive glue and solder. Such connectors act as electronic switches and instead of a wire a conductive thread is used in a smart garment; thus creating a fiber network. Humans are already conditioned using such connectors through their everyday experience with clothing. The interaction with a smart garment using such typical connectors makes the explanatory character of the functionality of a smart garment less difficult to understand for a wearer. A far more complicated issue deals with conductive yarns or thread. The + and – of a conductive thread needs to be separated when no conductive connection is necessary. Though, to allow this function the electronic fibers need to be insulated from each other. Their contact with the human skin is a main issue that needs to be addressed in the construction process. The skin is humid and can conduct electricity and as a result also create a short circuit. It can demolish the network the function is used for. The conductive thread is therefore best to be run on the surface of the textile or sandwiched between layers of textile or textile substrates. This constraint forms a new craft allowing to sustain the expertise of traditional textile design and to include nanotechnologies.

Power, Radiation, and the Environment

Besides the transport of energy throughout an item of clothing, the creation of energy is an important issue and strongly related to the way the wearer will interface and interact with the garment's functionalities. Energy for smart garments usually still comes from batteries, which are thrown away after their lifespan expires, causing huge environmental problems. Also, the life cycle of digitally enhanced devices in general is getting shorter and shorter. What happens with all the hazardous components and who is considering the energy consumption for the fabrication of such devices?

A battery's life is limited, thus the failure of a system depending on a battery is obvious. How can we avoid that a system, whose failure can be life threatening, collapses because of the lack of energy? If larger amounts of electricity are needed, research in solar power is probably the most advanced. Products like ScotteVest's (SeV) solar panels seem to be a first step towards using the surface of the garment that is constantly exposed to sunlight as a source of energy. ScotteVest's solar panels, trademarked PowerFLEX™, are flexible and can get wet, however, there are many 'do nots' in the user manual that make a widespread usage still questionable. An obvious and attractive energy source for intelligent clothing is the human body derived through movement or fluctuations in body temperature because of its characteristic of being less dependent on placement of the energy converter than solar or thermal energy. Today the energy that can be harvested through human kinetic energy based on movement or heat-exchange from the body can be measured in microwatts. It is, thus, too low to drive wearable technologies. Much research about sustainable energy consumption and the effects of wireless transmissions is still needed. Radiation and the effects of electro magnetic fields create a controversy amongst scientists and the industry, in particular when dealing with healthcare. Whether it is questionable to expose patients who are already weakened to additional pollution of electro magnetic frequency has yet to be understood. The effects of a monitoring system attached to a patient already wearing a pacemaker are one of the many issues that need to be tackled. As in the case of a network, the wearer needs to be able to switch off the power, to control the power consumption. Another option is to create a 'sleep' function like on a laptop; however, how can the wearer 'wake up' the system actively? What are the parameters for the automatic re-starting?

ERGONOMICS OF INTELLIGENT CLOTHING

To which extent do technical constraints impact the aesthetic and ability to wear the garments? An important difference makes the degree of body integration. The degree of intimacy is determined by personal preference or the functionality desired. The least integrated is a mobile device or handhelds like cell phones. These mobile devices can be attached to the garment, which serves as a container or a simple carrier. When such devices become accessories and are connected to the garment they become less mobile and more wearable. Tattoos are the most visible fully integrated visual display on the skin of the body. Less visible, but technically enhanced, are medical devices like implants. The degree of intimacy also depends on the context of use of the intelligent garment. Questions like how is the garment going to be used, when, by whom, for what, and why are asked to understand the use and therefore the human interaction with the garment.

Wearable technologies are essentially close to the skin, in particular through the use of sensors in the medical field. Textiles are more rigid than, for example, a hard cased microprocessor that is integrated into a garment. Gemperle, Kasabach, Stivoric et al. (1998) developed guidelines for wearability with their study 'Design of Wearability' and observed placement, form language, human movement, proximity, sizing, attachments, containment, weight, accessibility, sensory interaction, thermal, aesthetics, and long-term use. These considerations are a useful start for dealing with wearable technologies on our body and the construction of functional clothing. "A product that is wearable should have wearability. Wearability is defined as the interaction between the human body and the wearable object. Dynamic wearability extends that definition to include the human body in motion." (Gemperle, Kasabach, Stivoric et al., 1998, p. 1). Knight, Baber, Schwirtz et al. (2002) evaluated comfort across six dimensions being emotion, attachment, harm, perceived change, movement, and anxiety. Attachment, perceived change, and movement are sensed significantly stronger by the wearers, confirming the findings of Gemperle et al. (1998). A strong focus in the design of an intelligent garment needs to be placed on size, weight, weight distribution, placement, and attachment. The study also illustrates the importance of cognitive components like emotion and anxiety. In describing her research project 'whispers,' Schiphorst refers to these issues. "The research of whisper is based on wearable body architectures, extrapolated as small wearable devices, embedded within garments, worn close to skin: proximity creating resonance, contact and communication, body as carrier to device, device as devising the body" (Schiphorst, 2004, p. 1). Fashion designers, interaction designers, textile designers, and technologists are looking into innovative ways to integrate electronic components in the fashion design of the garments, while ensuring their wearability.

The Cut, Connectors, and Material

A wetsuit is closed with a zipper in the back of the suit with a long band to enable the wearer to reach it and use it. Thus the body ergonomics define the cut of the garment. Even though flexible antennas, batteries, and even microprocessors are developed the problem for their disappearance in a smart garment is an issue of the cut of the garment. Where can components such as these be hidden? How can a 'short-circuit be avoided? All this is an issue of the right cut, a major element in the design of the intelligent garment. The understanding of garment construction, the technical components, the network, the interaction, and the aesthetics is necessary for the success for the construction of an intelligent garment. Sizing looks at the volume of an embedded system; the dimensions of the microprocessor, the battery, and sensor, the antenna, etc determine the size of the garment. Layering can avoid short cuts and determines whether a separate layer for a specific sensor is necessary. The components and textiles have a specific weight and influence the garment's appearance. Draping creates the right shapes. The cut needs to allow for modular systems similar to Lego. Computer components need to be able to be easily exchanged due change of standards, failure, or simply the fact the garments needs to be washed. This requires modular systems including simple connection systems.

What are the current interactions with our garments? We need to button, zip, Velcro, pull over, slip into, and so much more. To understand our current interaction with garments it is necessary to develop a cut that integrates new, additional interactions. An example is the Burton Audex system with its textile switch interface on the sleeve to control an iPod in the jacket's interior pocket. The iPod is connected to the interface on the sleeve through a little microprocessor and conductive woven fabric in the sleeve of the jacket. The iPod can be removed and leaving a jacket that can easily be cleaned. Not only the technical functionality but also the extreme conditions were considered making it successful as an end-consumer product. In science and research, however, comfort and aesthetics are often not considered. Copper-coated conductive threads or fabrics might work ideally, conducting enough energy needed for the function the garment is promised to do. But, the textile is uncomfortable or even hazardous to wear or simply aesthetically not pleasing. The need for specialists in garment construction and textile design is apparent. Fashion designers often do not have a hands-on experience with the actual construction necessary for intelligent garments.

Aesthetics vs. Function

Building upon the considerations for wearability that define the relation between the human body and the wearable object, the question of aesthetics needs to be strongly examined. The cut is a major design element of architecture and construction of the garment. It defines the physical aesthetic of the garment—its appearance. The perceptions are many and range from cultural expectations to rules of human communication. The surface of a garment, "the cult of the body as an object of public display" (Warwick & Carvalho, 2004, p. 136), is once again noticeable in our culture. Thus a dynamic visual display on the garment calls for the wearer's control and need for confirmation of exposure if data. The functions of the visual display might change with its purpose and always need to be re-evaluated. Depending on the use of the garment either aesthetics or the functional components are more dominant. The wearer feels safe, protected from electromagnetic frequencies, which can also be considered a psychological function obvious. Social or cultural functions like ceremonial or religious strongly influence the look and feel of the garment. The functions of a specific intelligent garment are pre-defined in regards to their necessity. It is essential that technology and design work together to create garments that humans want to wear. And to understand the major differences in individual human needs. Another such cultural function of a garment is its association with a group. Groups of snowboarders sometimes tend to stencil their boards and their jackets. This is nothing new in our western culture but new technologies are used to create new applications by the users; not planned by those creating the technology.

THE FUTURE OF INTERACTION

"Not only did the garment impose a demeanor, it obliged me to live towards the exterior of the world" (Eco, 1976, p. 193). Commercially available successful interfaces are usable and intuitive. The use of durable technology is essential; wearable technology cannot break down. Users are conditioned that their PCs might fail but a cell phone never breaks down. Therefore the technology has to work; a wearer cannot just take off the garment and reboot it. Flexibility is a main characteristic of a textile interface and closely related to the comfort for the wearer and the cut the designer can choose. Washability is yet another major success factor; a garment needs to be able to be cleaned. Often it is enough to be able to hand-wash a garment that is not worn on the skin. Eleksen developed a smart fabric interface by pairing electronics with textile and calling it a fabric interface. Many of the claims Eleksen makes with its product are also the ones of other commercially successful textile interface manufacturers like Fibretronic. The interaction on the sleeve touchpads, however, is developed for right-handed people who are the majority of the population. Modular systems with plug-and-play connectors might seem to be a solution for many of the interaction problems. It also allows the quick removal of sensitive components that are essential for the washability of an intelligent garment. But, still a common standard needs to be developed to allow the use of components from different manufacturers in construction and use of an intelligent garment.

The human-garment interaction can either be passive when the wearer does not actively participate in the interaction or active, when a physical activation of the function by the wearer is necessary. Emotion-related physiological data from skin conductivity or skin temperature are an automatic, passive input. Such information needs to be translated in real-time and enable an immediate reaction to the current wearer-state. Active inputs can range from simple active mechanic input ranging from touch, pull, push, and so forth. to voice, eye tracking or facial recognition. The wearer needs to close the circuit through an active interaction. The Adidas_1, a running shoe by Adidas, first senses the runner's amount of compression in the heel, then the data is sent to a processor in the sole of the shoe, and the necessary adoption information is then sent to a motor-driving cable system in the shoe's muscle. Such a system requires the active usage of the shoe by the wearer but no actual interaction is required. The interaction, thus, is subtle and not obvious to the wearer. Picard's research in affective computing is particularly interesting when the user interaction is limited to inputs that are not using an input device or human hands. Picard (1998, p. 227) describes, "Wearables are computational devices that are worn as an article of clothing or jewelry. In particular, because of their potential for long-term intimate contact with you, wearables have a unique opportunity to become affective. An 'affective wearable' is a wearable system equipped with sensors and tools that enables recognition of its wearer's affective patterns." Very often the line blurs and passive inputs seem to become active when the wearer starts to influence the functionality of the intelligent garment.

The User Experience and its Considerations

The user experience with the intelligent garment is influenced by the ergonomics, the psychonomics, the interaction, and the context of use. The ergonomics describe the form of the wearable or an intelligent garment. The body ergonomics and the cut of the garment are the main influencers of wearability. The psychonomics refer to the perception from within and by the outside world. It is closely related to the aesthetics of the garment and social and cultural functions. The interaction, the feedback, is one of the most important factors to indicate that the communication took place and was successful or failed. The wearer always needs feedback in particular when the output, the dynamic visual element, is placed where the wearer cannot immediately see it. Depending on the use of the smart garment the interaction might change. The considerations for extreme sports situation with very low temperatures not only include the basic body ergonomics but also the human psychology. In a crisis situation, for example when hit by an avalanche, the wearer might react very differently to the same functionality the intelligent garment provides.

The comfort of an embedded system is experienced strongly by the wearer. A sensor that reads biometric data from the skin needs to be close to the body but also comfortable to wear. Bodymedia is promoting its 3.0 armband to be worn like a Band-Aid close to the skin without the need of an armband bulkier than a watch. The output is shown on a computer screen, the information transferred wirelessly. However, an immediate display of the data onto the wearers garment would allow for a self-enclosed body area network with and immediate ability for changes. The applications and the areas of display on the surface need to be considered wisely. Such a system not only requires comfort, it also needs to be safe. The wearer needs to feel in control and as a result safe; in particular in wellness, rehabilitation, and chronic diseases where the wearer is much aware of what is happening. Such a system requires special attention regarding the interaction possibilities. A successful intelligent garment interaction considers all factors: human, environment, and machine. The guideline for design considerations for an intelligent garment in Table 2 refers to the factors that influence the wearer's experience.

Table 2. Design considerations

Factors	Considerations
Body ergonomics	wearability and overall comfort (placement, form language, human movement, proximity, sizing, attachments, weight, accessibility, heat, material, cut)
Psychonomics	perceptions by the wearer and the environment, aesthetics, psychological function
Interaction	interface with the system (e.g. inputs), practicality of daily use (e.g., washing/ cleaning)
Context of use	functions (social, cultural, physical), environment, wearer

Applications

The applications and functions of conductive fibers are countless. Clothing with embedded technologies is evident in sport, work wear, healthcare and rehabilitation, rescue services, and elderly care and prevention. The consumer interest in functional wear in particular in sports is steadily increasing and the monetary threshold is rising. Smart fabrics like Goretex have set the price for 'extreme' sports wear with enhanced functionality high for the consumer market. It is easier now for Nike with its Nike+ or Burton's Audex system to reach the price levels needed to sustain the development of these products with enhanced embedded technology. Other industries will follow when solutions for the issues around usability, the interaction, and the aesthetics are honestly confronted.

The dynamic real-time output on the surface of the intelligent garment is limitless and depends on the applications and functions that are needed. A visual output is limiting and its suitability needs to be examined. Obvious commercial applications like wearable advertising, safety information or the visual output of body area network data in healthcare are relevant future applications.

CONCLUSION

"Technology has enabled a great deal of personalization in fashion" (Seymour, 2004, p. 534) in intelligent garments. Intelligent garments and wearable devices are very personal. The personal relation to a garment is very sound because of its many perceptions. They are stronger than those associated with mobile devices that can be hidden in a pocket. The human-garment relationship is therefore an important consideration. The multidisciplinary team that designs and constructs the intelligent garment needs to understand the wearer's needs, the context of use, and the relevant factors of the wearing experience. Besides the obvious trades like textiles, electronics, and fashion design the knowledge and expertise of psychologists, philosophers, representatives from the medical professions, anthropologists, and so on are essential to create successful and sustainable intelligent garments. 1) The user experience is influenced by the wearer's interaction with the intelligent garment. 2) Only the thoughtful construction of wearable systems allow for a great user experience and useful functionality. Consequently the design of user experience comprises all main design factors to enable successful use of an intelligent garment by an individual.

REFERENCES

Anttila, A., Ribak, A., Cereijo-Roibás, A., Seymour, S., Svanteson, S., Weiss, S. et al. (2004). Mobile communication versus pervasive communications: The role of handhelds. In Brewster, S. & Dunlop, M. (Eds.), *Mobile human-computer interaction, MobileHCI 2004* (pp. 531–535). Berlin, Germany: Springer-Verlag.

Baurley, S. (2005). Interaction design in smart textile clothing and applications. In Tao, X. (Ed.), *Wearable electronics and photonics* (pp. 223–243). Cambride, UK: Woodhead Publishing Ltd.

Berzowska, J. (2005). Electronic textiles: Wearable computers, reactive fashion, and soft computation. In Jeffries, J. (Ed.), *Textile, the journal of cloth & culture, digital dialogues 2: Textiles and technology, 3*(1), 2-19. Biggleswade, UK: Berg Publishers.

Braddock, S. E., & O'Mahony, M. (2006). *Techno textiles 2. Revolutionary fabrics for fashion and design*. New York, NY: Thames & Hudson.

Eco, U. (1990). *Travels in hyperreality.* Fort Washington, PA: Harvest Books.

Farren, A. & Hutchison, A. (2004). Digital clothes: Active, dynamic, and virtual textiles and garments. In Jeffries, J. (Ed.), *Textile, The Journal of Cloth & Culture, Digital Dialogues 2: Textiles and Technology, 2*(3), 290-306. Biggleswade, UK: Berg Publishers.

Gemperle, F., Kasabach, C., Stivoric, J., Bauer, M., & Martin, R. (1998). *Design for wearability.* Pittsburgh, PA: Carnegie Mellon University. Retrieved April 22, 2002, from http://www.ices.cmu.edu/design/wearability

Knight, J. F., Baber, C., Schwirtz, A., & Bristow H. W. (2002). The comfort assessment of wearable computers. In *6th International Symposium on Wearable Computers (ISWC 2002)* (pp. 65-74). Las Alamitos, CA: IEEE Computer Society.

Koncar, V., Deflin, E., & Weill, A. (2005). Communication apparel and optical fibre fabric display. In Tao, X. (Ed.), *Wearable Electronics and Photonics* (pp. 155-176). Cambride, UK: Woodhead Publishing Ltd.

McCann, J., Hurford, R., & Martin, A. (2005). A design process for the development of innovative smart clothing that addresses end-user needs from technical, functional, aesthetic and cultural view points. In *the 9th IEEE International Symposium on Wearable Computers* (pp. 70 – 77). Las Alamitos, CA: IEEE Computer Society.

McQuaid, M. (Ed.). (2005). *Extreme textiles. Designing for high performance*. New York, NY: Princeton Architectural Press.

Peter, C., Ebert, E, & Beikirch, H. (2005). A wearable multi-sensor system for mobile acquisition of emotion-related physiological data. In J. Tao, T. Tan, & R. W.Picard (Eds.), *Affective computing and intelligent interaction. First international conference*, *ACII 2005* (pp. 691-698). Berlin, Germany: Springer-Verlag.

Picard, R. W. (1997). *Affective computing.* Cambridge, MA: MIT Press.

Raffle, H., Ishi H., & Tichenor, J. (2004). Super cilia skin: A textural interface. In Jeffries, J. (Ed.), *Textile, the journal of cloth & culture, digital dialogues 2: Textiles and technology*, *2*(3) (pp. 328-347). Biggleswade, UK: Berg Publishers.

Rosson, M. B., & Carroll, J. M. (2002). *Usability engineering. Scenario-based development of human-ccomputer interaction*. San Francisco: Morgan Kaufmann Publishers.

Schiphorst, T. (2004). *Soft, softer and sofly: [whispering] between the lines.* Surrey, Canada: Simon Fraser University. Retrieved August 16, 2006, from http://whisper.surrey.sfu.ca/PDF/whispering-between-lines.pdf

Seymour, S. (2004). The complexity of collaboration in wearable computing. In Tralla, M. (Ed.), *RAM3. Reclaiming cultural territory in new media.* (pp. 13-16). Tallinn, Estonia: RAM.

Warwick, A., & Cavallaro, D. (2001). *Fashioning the frame. Boundries, dress and the body.* Oxford, UK: Berg.

KEY TERMS

Body Area Network (BAN): Connects independent nodes dispersed in a smart garment. For example implanted medical devices and on-body sensors are connected wirelessly with monitoring tools to provide patient health data in real-time.

Conductive Threads, Electronic Fibers, or Conductive Yarns: Textile-based yarns that conduct electricity. The most rudimentary form is copper-coated yarn.

Connectors in Smart Garments: Are conductive fasteners such as metal-based snaps or zippers that function like a switch in an electronic circuit.

Electronics Textiles, Smart Fabrics, or Conductive Textiles: Result from the integration of technology into a textile. A textile-based circuit board is achieved through incorporating conductive yarns into the fabric.

Fiber Network: An electrically conductive network within the fabric and works with all technical components like inputs, outputs, microprocessors, and communication networks.

Intelligent Garments: Functional clothing constructed with electronic textiles or smart fabrics and wearable technologies to enable various functionalities.

Wearables: Objects with embedded technologies that are wearable and may have wireless communication capabilities.

Wearable Technologies: Technologies or systems that are embedded into intelligent garments or wearable objects. Components are inputs, outputs, microprocessors, and communication networks.

Chapter XII
Context as a Necessity in Mobile Applications

Eleni Christopoulou
University of Patras & Ionian University, Greece

ABSTRACT

This chapter presents how the use of context can support user interaction in mobile applications. It argues that context in mobile applications can be used not only for locating users and providing them with suitable information, but also for supporting the system's selection of appropriate interaction techniques and providing users with a tool necessary for composing and creating their own mobile applications. Thus, the target of this chapter is to demonstrate that the use of context in mobile applications is a necessity. It will focus on the current trend of modeling devices, services and context in a formal way, like ontologies, and will present an ontology-based context model.

INTRODUCTION

The future of computer science was marked by Weiser's vision (Weiser, 1991), who introduced the term ubiquitous computing (ubicomp) by defining a technology that can be seamlessly integrated into the everyday environment and aid people in their everyday activities. A few years later, the European Union, aiming to promote "human-centered computing," presented the concept of ambient intelligence (AmI) (ISTAG, 2001), which involves a seamless environment of computing, advanced networking technology and specific interfaces. So, technology becomes embedded in everyday objects such as furniture, clothes, vehicles, roads, and smart materials, providing people with the tools and processes that are necessary in order to achieve a more relaxing interaction with their environment.

Several industry leaders, like Philips and Microsoft, have turned to the design of ubicomp applications with a focus on smart home applications. However, people nowadays are constantly on the move, travel a lot, and choose to live in remote or mobile environments. In the near future, each person will be "continually interacting with hundreds of nearby wirelessly connected computers" (Weiser, 1993). Therefore, the need for mobile applications is now more evident than ever.

Recent years have seen a great breakthrough occur in the appearance of mobile phones. Initially they were used as simple telephone devices. Today, mobiles have evolved into much more than that. Although the majority of people still use mobile phones as communication devices, an increasing number of users have begun to appreciate their potential as information devices. People use their smart mobile phones to view their e-mails, watch the news, browse the Web, and so forth. Eventually, mobile phones and other mobile handheld devices became an integral part of our daily routine.

Both scientists and designers of ubicomp applications have realized that the mobile phone could be considered as one of the first AmI artefacts to appear. As mobile phones are becoming more powerful and smarter this fact is increasingly proven true. Thus, scientists wanting to take advantage of the emerging technology have implemented a great number of mobile applications that enable human-computer interaction through the use of handheld devices like mobile phones or personal digital assistants (PDAs). Such applications include visitor guides for cities and museums, car navigation systems, assistant systems for conference participants, shopping assistants and even wearable applications.

A closer examination of mobile applications shows that most of them are location-aware systems. Specifically, tourist guides are based on users' location in order to supply more information on the city attraction closer to them or the museum exhibit they are seeing. Nevertheless, recent years have seen many mobile applications trying to exploit information that characterizes the current situation of users, places and objects in order to improve the services provided. Thus, context-aware mobile applications have come to light.

Even though significant efforts have been devoted to research methods and models for capturing, representing, interpreting, and exploiting context information, we are still not close to enabling an implicit and intuitive awareness of context, nor efficient adaptation to behavior at the standards of human communication practice. Most of the current context-aware systems have been built in an ad-hoc approach, deeply affected by the underlying technology infrastructure utilized to capture the context (Dey, 2001). To ease the development of context-aware ubicomp and mobile applications it is necessary to provide universal models and mechanisms to manage context.

Designing interactions among users and devices, as well as among devices themselves, is critical in mobile applications. Multiplicity of devices and services calls for systems that can provide various interaction techniques and the ability to switch to the most suitable one according to the user's needs and desires. Context information can be a decisive factor in mobile applications in terms of selecting the appropriate interaction technique.

Another inadequacy of current mobile systems is that they are not efficiently adaptable to the user's needs. The majority of ubicomp and mobile applications try to incorporate the users' profile and desires into the system's infrastructure either manually or automatically observing their habits and history. According to our perspective, the key point is to give them the ability to create their own mobile applications instead of just customizing the ones provided.

The target of this chapter is to present the use of context in context-aware ubicomp and mobile applications and to focus on the current trend of modeling devices, services and context in a formal way (like ontologies). Our main objective is to show that context in mobile applications can be used not only for locating users and providing them with suitable information, but also for supporting the system's selection of appropriate interaction techniques and for providing them with a tool necessary for composing and creating their own mobile applications.

In the background section, which follows, we define the term context and present how context is modeled and used in various mobile applications focusing on ontology-based context models. In the subsequent sections we present our perspective of context, an ontology-based context model for mobile applications as well as the way in which human-computer interaction can be supported by the use of context. The Future section embraces our ideas of what the future of human-computer interaction in mobile applications can bring by taking context into account. Finally we conclude with some prominent remarks.

BACKGROUND

What is Context

The term "context-aware" was first introduced by Schilit and Theimer (1994), who defined context as "the location and identities of nearby people and objects, and changes to those objects." Schilit, Adams, and Want (1994) defined context as "the constantly changing execution environment" and they classified context into computing environment, user environment, and physical environment. Schmidt (2000) also considered situational context, such as the location or the state of a device, and defined context as knowledge about the state of the user and device, including surroundings, situation and tasks and pointing out the fact that context is more than location.

An interesting theoretical framework has been proposed by Dix et al. (2000), regarding the notions of space and location as constituent aspects of context. According to this framework context is decomposed into four dimensions, which complement and interact with each other. These dimensions are: system, infrastructure, domain, and physical context.

One of the most complete definitions for context was given by Dey and Abowd (2000); according to them context is "any information that can be used to characterize the situation of an entity. An entity should be treated as anything relevant to the interaction between a user and an application, such as a person, a place, or an object, including the user and the application themselves."

When studying the evolution of the term "context" one notices that the meaning of the term has changed following the advances in context-aware applications and the accumulation of experience in them. Initially the term "context" was equivalent to the location and identity of users and objects. Very soon, though, the term expanded to include a more refined view of the environment assuming either three major components; computing, user and physical environment, or four major dimensions; system, infrastructure, domain, and physical context. The term did not include the concept of interaction between a user and an application until Dey and Abowd (2000). This definition is probably at present the most dominant one in the area.

Context Modeling in Context-Aware Applications

A number of informal and formal context models have been proposed in various systems; the survey of context models presented in Strang and Linnhoff-Popien (2004) classifies them by the scheme of data structures. In Partridge, Begole and Bellotti (2005) the three types of contextual models, which are evaluated, are environmental, personal, and group contextual model.

Among systems with informal context models, Context Toolkit (Dey, Salber & Abowd, 2001) represents context in the form of attribute-value tuples, and Cooltown (Kindberg et al., 2002) proposed a Web-based model for context in which each object has a corresponding Web description. Both ER and UML models are used for the representation of formal context models in Henricksen, Indulska, and Rakotonirainy (2002).The context modeling language is used in Henricksen and Indulska (2006) in order to capture user activities, associations between users and communication channels and devices and locations of users and devices.

Truong, Abowd and Brotherton (2001) point out that the minimal set of issues required to be addressed when designing and using applications are: who the users are, what is captured and accessed, when and where it occurs, and how this is performed. Designers of mobile applications should also take these issues into account. Similar to this approach Jang, Ko and Woo (2005) proposed a unified model in XML that represents user-centric contextual information in terms of 5W1H (who, what, where, when, how, and why) and can enable sensor, user, and service to differently generate or exploit a defined 5W1H-semantic structure.

Given that ontologies are a promising instrument to specify concepts and their interrelations (Gruber, 1993; Uschold & Gruninger1996), they can provide a uniform way for specifying a context model's core concepts as well as an arbitrary amount of subconcepts and facts, altogether enabling contextual knowledge sharing and reuse in a Ubicomp system (De Bruijn, 2003). Ontologies are developed to provide a machine-processable semantics of information sources that can be communicated between different agents (software and humans). A commonly accepted definition of the term ontology was presented by Gruber (1993) and stated that "an ontology is a formal, explicit specification of a shared conceptualization." A "conceptualization" refers to an abstract model of some phenomenon in the world which identifies the relevant concepts of that phenomenon; "explicit" means that the type of concepts used and the constraints on their use are explicitly defined and "formal" refers to the fact that the ontology should be machine readable. Several research groups have presented ontology-based models of context and used them in ubicomp and mobile applications. We will proceed to briefly describe the most representative ones.

In the Smart Spaces framework GAIA (Ranganathan & Campbell, 2003) an infrastructure that supports the gathering of context information from different sensors and the delivery of appropriate context information to ubicomp applications is presented; context is represented as first-order predicates written in DAML+OIL. The context ontology language (Strang, Linnhoff-Popien & Frank, 2003) is based on the aspect-scale-context information model. Context information is attached to a particular aspect and scale and quality metadata are associated with information via quality properties. This contextual knowledge is evaluated using ontology reasoners, like F-Logic and OntoBroker.

Wang, Gu, Zhang et al. (2004) created an upper ontology, the CONON context ontology, which captures general features of basic contextual entities, a collection of domain specific ontologies and their features in each subdomain. An emerging and promising context modeling approach based on ontologies is the COBRA-ONT (Chen, Finin & Joshi, 2004). The CoBrA system provides a set of OWL ontologies developed for modeling physical locations, devices, temporal concepts, privacy requirements and several other kinds of objects within ubicomp environments.

Korpipää, Häkkilä, Kela et al. (2004) present a context ontology that consists of two parts: structures and vocabularies. Context ontology, with the enhanced vocabulary model, is utilized to offer scalable representation and easy navigation of context as well as action information in the user interface. A rule model is also used to allow systematic management and presentation of context-action rules in the user interface. The objective of this work is to achieve personalization in mobile device applications based on this context ontology.

Although each research group follows a different approach for using ontologies in modeling and managing context in ubicomp and mobile applications, it has been acknowledged by the majority of researchers (Biegel & Cahill, 2004; Dey et al., 2001; Ranganathan & Campbell, 2003) that it is a necessity to decouple the process of context acquisition and interpretation from its actual use, by introducing a consistent, reliable and secure context framework which can facilitate the development of context-aware applications.

Context Utilisation in Mobile Applications

In context-aware mobile applications location is the most commonly used variable in context recognition as it is relatively easy to detect. Thus, a lot of location-aware mobile systems have been designed, such as shopping assistants (Bohnengerger, Jameson, Kruger et al., 2002) and guides in a city (Davies, Cheverst, Mitchell et al., 2001) or campus area (Burrell, Gay, Kubo et al., 2002). Many location-aware mobile applications are used in museum environments; a survey is presented in (Raptis, Tselios & Avouris, 2005). In the survey of Chen and Kotz (2000) it is evident that most of the context-aware mobile systems are based on location, although some other variables of context like time, user's activity and proximity to other objects or users are taken into consideration.

User activity is much more difficult to identify than location, but some aspects of this activity can be detected by placing sensors in the environment. Advanced context-aware applications using activity context information have been put into practice for a specific smart environment (Abowd, Bobick, Essa et el., 2002). The concept of activity zones (Koile, Tollmar, Demirdjian et al., 2003) focuses on location, defines regions in which similar daily human activities take place, and attempts to extract users' activity information from their location.

Sensor data can be used to recognize the usage situation based on illumination, temperature, noise level, and device movements, as described for mobile phones in Gellersen, Schmidt and Beigl (2002) and PDA in Hinkley, Pierce, Sinclair et al.(2000), where it is suggested that contextual information can be used for ring tone settings and screen layout adaptation. The mobile device can observe the user's behavior and learn to adapt to a manner that is perceived to be useful at a certain location as was the case with the comMotion system (Marmasse & Schmandt, 2000).

Sadi and Maes (2005) propose a system that can make adaptive decisions based on the context of interaction in order to modulate the information presented to the user or to carry out semantic transformation on the data, like converting text to speech for an audio device. CASIS (Leong, Kobayashi, Koshizuka et al., 2005) is a natural language interface for controlling devices in intelligent environments that uses context in order to deal with ambiguity in speech recognition systems. In Häkkilä and Mäntyjärvi (2005) context information is used in order to improve collaboration in mobile communication by supplying relevant information to the cooperating parties, one being a mobile terminal user and the other either another person, group of people, or a mobile service provider.

Perils of Context-Awareness

The promise and purpose of context-awareness is to allow computing systems to take action autonomously; enable systems to sense the situation and act appropriately. Many researchers, though, are skeptical and concerned because of the problems that emerge from context-awareness.

A main issue regarding context-aware computing is the fear that control may be taken away from the user (Barkhuus & Dey 2003). Experience has shown that users are still hesitant to adopt context-aware systems, as their proactiveness is not always desired. Another aspect of this problem is that users often have difficulties when presented with adaptive interfaces.

Apart from control issues, privacy and security issues arise. The main parameters of context are user location and activity, which users consider as part of their privacy. Users are especially reluctant to exploit context-aware systems, when they know that private information may be disclosed to others (Christensen et al., 2006).

Even recent research projects suffer from difficulties in automated context fetching; in order to overcome this, the user is asked to provide context manually. Studies have shown that users are not willing to do much in order to provide context and context that depends on manual user actions is probably unreliable (Christensen et al., 2006). Additionally, systems that ask from users to supply context fail, as this affects the user's experience and diminishes his benefit from the system.

Practice has shown that there is a gap between how people understand context and what systems consider as context. The environment in which people live and work is very complex; the ability to recognize the context and determine the appropriate action requires considerable intelligence. Skeptics (Erickson, 2002) believe that a context-aware system is not possible to decide with certainty which actions the user may want to be executed; as the human context is inaccessible to sensors, we cannot model it with certainty. They, also, argue whether a context-aware system can be developed to be so robust that it will rarely fail, as ambiguous and uncertain scenarios will always occur and even for simple operations exceptions may exist. A commonly applied solution is to add more and more rules to support the decision making process; unfortunately this may lead to large and complex systems that are difficult to understand and use.

An issue that several researchers bring forward (Bardram, Hansen, Mogensen et al., 2006) is that context-aware applications are based on context information that may be imperfect. The ambiguity over the context soundness arises due to the speed at which the context information changes and the accuracy and reliability of the producers of the context, like sensors.

It is a challenge for context-aware systems to handle context, that may be non accurate or ambiguous, in an appropriate manner. As Moran and Dourish (2001) stated, more information is not necessarily more helpful; context information is useful only when it can be usefully interpreted.

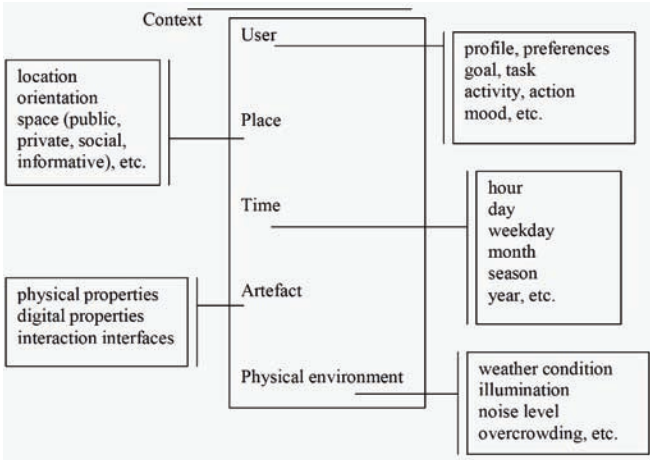
WHAT IS CONTEXT FOR MOBILE APPLICATIONS?

Considering the use of context in the mobile applications discussed in the background section, we may conclude that, for these applications, context is almost synonymous to location and, specifically, to user location. However, context is quite more than just that. In this section, we will present our perspective on the parameters of context that are necessary for mobile applications. In order to figure out these parameters we have to identify the concepts that constitute the environment in which mobile applications exist. The primary concepts are indubitably people, places, time, objects and physical environment.

A mobile application is context-aware if it uses context to provide relevant information to users or to enable services for them; relevancy depends on a user's current task and profile. The user context issue has been addressed by many researchers of context-awareness (Crowley, Coutaz, Rey et al., 2002; Schimdt, 2002). However, the key for context-aware mobile applications is to capture user activity and preferences. Apart from knowing who the users are and where they are, we need to identify what they are doing, when they are doing it, and which object they focus on. In the background section we mentioned that, until now, most mobile applications determine user activity by their location; it is apparent, however, that a more elaborate model is necessary for representing this activity. Stahl (2006) proposes a model that represents a user's goals, activities and actions; he suggests that the distinction between an activity and an action lies in the fact that an activity takes a time span, while actions occur instantaneously. The system can define user activity by taking into account various sensed parameters like location, time, and the object that they use. For example, when a user opens the front door he is thought to be either entering or leaving the house, when the bed is occupied and the television is turned on he is watching a movie, but when the television is turned off he is probably sleeping. User preferences are also very important for context-aware mobile applications, but it is difficult for the system to define them. Users have to incorporate their preferences into the application on their own, although the system can also gather information from the interaction with them in order to acquire experience based on history. By exploiting system experience the application may also infer a user's mood, a factor that cannot be measured by any sensor.

In order to identify user location various technologies are being used. In outdoors applications, and depending on the mobile devices that are used, satellite supported technologies, like GPS, or network supported cell information, like GSM, IMTS, WLAN, are applied. Indoors applications use RFID, IrDA and Bluetooth technologies in order to estimate the users' position in space. Although location is the determining factor in identifying where users are, orientation is also a very important parameter; the system has to know what users are looking at or where they are going to. However, in order to efficiently exploit the information on user location and orientation, the mobile application needs to have a representation of the layout of the place in which users are. Spaces can be classified into the following types: public,

Figure 1. Context in mobile applications



private, an area in which restrictions may apply, transient, places where people do not congregate easily or frequently, like hallways and corridors, social, public places where people arrange to meet, like coffee shops, informative, places that are used for public announcements (Mitchell, Race & Suggitt, 2006). Additionally a space can also be divided into districts, for example a home may have a living room, kitchen and bedroom, while a museum could have ancient Greek, paintings and modern art sections, as well into zones, such as lower left, upper left, and so forth.

Time is another significant parameter of context as it can play an important role in order to extract information on user activity; for example if it is early in the morning and the front door is opening the user is probably leaving the house, not entering it. Time can be used in various forms such as hour (daytime), night, day, weekday, week, month, season and year.

The objects that are used in mobile applications are the most crucial context sources. In mobile applications the user can use mobile devices, like mobile phones and PDAs and objects that are enhanced with computing and communication abilities (AmI artefacts). Sensors attached to artefacts provide applications with information about what the user is utilizing. However, this is not the most important parameter of context sensed by the artefacts. In order to present the user with the requested information in the best possible form, the system has to know the physical properties of the artefact that will be used, for example the display size of the artefact is determinant for the modulation of information. Additionally, the types of interaction interfaces that an artefact provides to the user need to be modeled; the system has to know if an artefact can be handled by both speech and touch techniques or if a mobile phone can vibrate. Apart from the physical properties of an artefact, the system must know how it is designed. A table with only one weight sensor in the centre cannot provide to the application information on whether an object is at its edge; thus the system has to know the number of each artefact's sensors and their position in order to gradate context information with a level of certainty. Based on information on the artefact's physical properties and capabilities, the system can extract information on the services that they can provide to the user; this is considered to be the most crucial context information related to artefacts. The application has to know if a printer can print both black-and-white and color text or if it can supply free maps and guidelines to a user that is close enough to a city's info center.

Finally, context from the physical environment may include current weather conditions, illumination, noise level, overcrowding. Taking into account the illumination of a room the application may decide to turn on an additional light when a user is reading a book or, if a user is in a noisy public space, the system may decide to vibrate his mobile phone when he has a call.

We selected to model the parameters of context illustrated in Figure 1 creating an ontology and taking into account the acknowledgement, shared by the majority of researchers (Biegel & Cahill, 2004; Dey et al., 2001; Ranganathan & Campbell, 2003), that it is a necessity to decouple the process of context acquisition and interpretation from its actual use. In the next section the details of this ontology-based context model are discussed.

AN ONTOLOGY-BASED CONTEXT MODEL FOR MOBILE APPLICATIONS

The key idea behind the proposed context model is that artefacts of AmI environments can be treated as components of a context-aware mobile application and users can compose such applications by creating associations between these components. In the proposed system, artefacts are considered as context providers. They allow users to access context in a high-level abstracted form and they inform other application's artefacts so that context can be used according to the application needs. Users are able to establish associations between the artefacts based on the context that they provide; keep in mind that services enabled by artefacts are provided as context. Thus defining the behavior of the application that they create, they can also denote their preferences, needs and desires to the system.

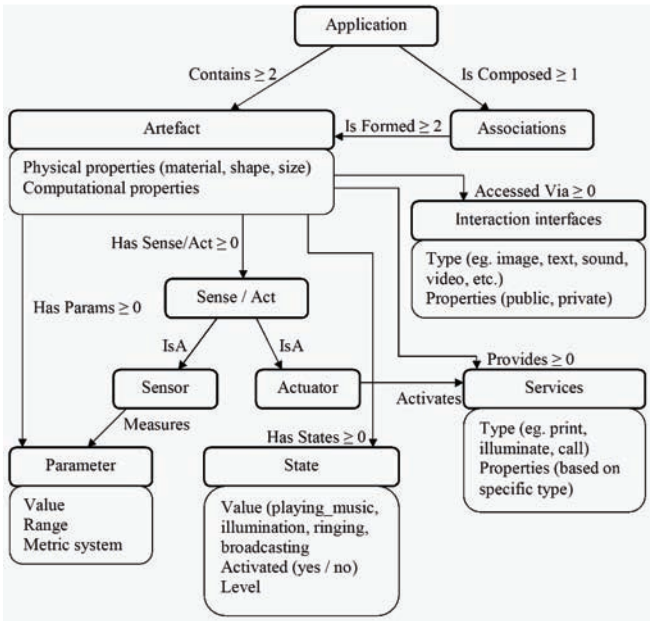
The set of sensors attached to an artefact measure various parameters such as location, time, temperature, proximity, motion, and so forth; the raw data given by its sensors is the artefact's low level context. As the output of different sensors that measure the same artefact parameter may differ, for example sensors may use different metric system, it is necessary to interpret the sensors' output into higher level context information. Aggregation of context is also possible meaning that semantically richer information may be derived based on the fusion of several measurements that come from different homogeneous or heterogeneous sensors. Thus, an artefact based on its own experience and use has two different levels of context; the low level which represents information acquired from its own sensors and the high level that is an interpretation of its low level context information. Additionally, an artefact can get context information from the other artefacts; this context can be considered as information from a "third-person experience."

When a user interacts and uses an artefact it affects its state; for example turning on the television sets it in a different state. An artefact may decide to activate a response based on both a user's desires and these states; for example when the user's PDA perceives that it is close to a specific painting in a museum, it will seek information about this painting. Such decisions may be based on the artefact's local context or may require context from other artefacts. The low and high level context, their interpretation and the local and global decision-making rules can be encoded in an ontology.

The ontology that we propose to represent the context of mobile applications is based on the GAS Ontology (Christopoulou & Kameas, 2005). This ontology is divided into two layers: a common one that contains the description of the basic concepts of context-aware applications and their inter-relations representing the common language among artefacts and a private one that represents an artefact's own description as well as the new "knowledge or experience" acquired from its use.

The common ontology, depicted in Figure 2, defines the basic concepts of a context-aware application; such an application consists of a number of artefacts and their associations. The concept of artefact is described by its physical properties and its communication and computational capabilities; the fact that an artefact has a number of sensors and actuators attached is also defined in our ontology. Through the sensors an artefact can perceive a set of parameters based on which the state of the

Figure 2. The common ontology



artefact is defined; an artefact may also need these parameters in order to sense its interactions with other artefacts as well as with the user. Artefacts may provide various services to the environment, for example a printer provides the print service, a lamp provides illumination and a phone the call service; these services are activated either by the user or by other artefacts using the actuators attached to artefacts. The interaction interfaces via which artefacts may be accessed are also defined in our ontology in order to enable the selection of the appropriate one.

We have decided that each parameter of context in our context-aware mobile applications, for example user, space, time and physical environment, is represented as an application's artefact. For instance, the notion of time is integrated into such applications only if a watch or a clock may provide this context as a service. The necessary information about the users that interact with such applications may be provided by the users' mobile phone or PDA. The services provided by such artefacts may be regarded as context; for instance the information that a thermometer provides is context related to the weather and we consider that the thermometer provides a temperature service. So, based on the concepts of context and their subcategories as presented in Figure 1, we have designed a service classification.

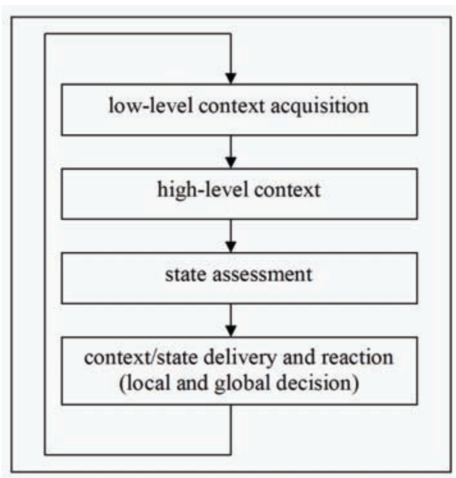
The common ontology represents an abstract form of the concepts represented, especially of the context parameters, as more detailed descriptions are stored into each artefact's private ontology. For instance, the private ontology of an artefact that represents a house contains a full description of the different areas in a house as well as their types and their relations.

The question that arises is where should these ontologies be stored? The system's infrastructure is responsible for answering this question. For a centralized system the common ontology as well as all the artefacts' ontologies can be stored in a central base. However, the majority of context-aware mobile applications are based on ad-hoc or p2p systems. Therefore, we propose that each artefact should store the common ontology and its private one itself; although when an artefact has limited memory resources its private ontology could be stored somewhere else. Another issue is where should place, time, environment and user ontologies be stored? The artefact that measures time, for example a clock, is responsible to store the time ontology; similarly there is an artefact for the environmental context. The place ontology can be stored either in a specific artefact that represents the space, for example an info kiosk in the entrance of a museum, or in the digital representation of the space managed by the application, for example the context from the sensors located in a room should be handled by the system through the use of the e-room and stored in an artefact with sufficient memory and computational capabilities. The user ontology in a similar way to the place ontology can be stored either in the user's mobile phone or in a digital self. These ontologies could also be stored in a web server in order to be accessible from artefacts.

The basic goal of the proposed ontology-based context model is to support a context management process, presented in Figure 3, based on a set of rules that determine the way in which a decision is made and are applied to existing knowledge represented by this ontology. The rules that can be applied during such a process belong to the following categories: rules for an artefact's state assessment that define the artefact's state based on its low and high level context, rules for local decisions which exploit an artefact's knowledge only in order to decide the artefact's reaction (like the request or the provision of a service) and finally rules for global decisions that take into account various artefacts' states and their possible reactions in order to preserve a global state defined by the user (Christopoulou, Goumopoulos & Kameas, 2005).

The ontology that is the core of the described context-management process was initially developed in the extrovert-Gadgets (eGadgets) project (http://www.extrovert-gadgets.net). In the e-Gadgets project our target was to design and develop an architectural framework (the Gadgetware Architectural Style—GAS) that would support the composition of ubicomp applications from everyday physical objects enhanced with sensing, acting, processing and communication abilities. In this project we implemented the GAS Ontology (Christopoulou & Kameas, 2005), which served the purpose of describing the semantics of the basic concepts of a ubicomp environment and defining their inter-relations. The basic goal of this ontology was to provide a common language for the communication and collaboration among the heterogeneous devices that constitute these environments; it also supported a service discovery mechanism necessary for that ubicomp environment. Already, at this early stage, we had decided on issues like how this ontology would be stored in each artefact, by dividing it into two layers, and a module had been implemented, which was responsible for managing and updating this ontology.

Figure 3. Context-management process



This work evolved in the PLANTS project (http://plants.edenproject.com) that aimed to enable the development of synergistic, scalable mixed communities of communicating artefacts and plants (Goumopoulos, Christopoulou, Drossos et al., 2004). In this project we extended the concept of "context" in order to allow for the inclusion of plants as components of our ubicomp applications, by attaching sensors to them that provided information regarding the plants' state. The ontology that was inherited from the e-Gadgets project was extended and refined in order to include all the parameters of context that were identified as necessary for our applications. The ontology-based context model and the context-management process, presented in Figure 3, were defined at that stage. Experience showed that our system managed to decouple the process of context acquisition and interpretation from its actual use. Our context-management process is based on a set of rules that define the state of each artefact or plant in an application. Based on these rules and their state, each artefact determines its local decisions; the set of rules on various artefacts determine global decisions made by the whole application. These rules are defined by the users themselves via a graphical user interface. Each artefact stores its ontology as well as its rule base as defined by the user; the decision-making process, part of the context-management process, is supported by an inference engine. Experience has shown that users could easily define their own applications, denoting the rules that govern both each artefact and the whole application; the fact that the reasoning process permits user-defined rules that can be dynamically updated was another positive point. A drawback of our system is that the inference engine, which was used required significant memory that was not always available; a workaround to this problem was to host the inference engine in an artefact with the required capabilities. Details on the design and implementation of this system as well as a case study of an application in the e-health domain and an evaluation of the outcome are presented in Christopoulou et al. (2005).

HOW CONTEXT CAN SUPPORT USER INTERACTION IN MOBILE APPLICATIONS

Recalling the use of context in mobile applications presented in the background section, we reach the conclusion that context has not been adequately exploited so far in order to support human-computer interaction. In this section we will present how our ontology-based context model enables the use of context in order to assist human-computer interaction in mobile applications and to achieve the selection of the appropriate interaction technique.

The goal of context in computing environments is to improve interaction between users and applications. This can be achieved by exploiting context, which works like implicit commands and enables applications to react to users or surroundings without the users' explicit commands (Schmidt, 2000). Context can also be used to interpret explicit acts, making interaction much more efficient. Thus, context-aware computing completely redefines the basic notions of interface and interaction.

The future of human computer interaction is going further than WIMP (Windows Icons Menus Pointing) interfaces. Jones and Marsden (2005) present various mobile interaction techniques that are trying to better exploit a user's capabilities like auditory (hearing) and haptic (touch and movement sensing) abilities as well as gestural skills, such as the expressive movements users can make with their hands or heads. More senses (vision, hearing, touch) and more means of expression (gestures, facial expression, eye movement and speech) are involved in human-computer interaction. A comparable analysis of mobile interaction techniques is presented in (Ballagas, Borchers, Rohs et al., 2006).

Rukzio et al. (2006) conclude from their experimental comparison of touching, pointing and scanning interaction techniques that users tend to switch to a specific physical mobile interaction technique dependent on location, activity and motivation; for example when a user is close enough to an artefact he prefers to touch it, otherwise he has no motivation for any physical effort. Thus, mobile systems have to provide multi-modal interfaces so that users can select the most suitable technique based on their context.

The ontology-based context model that we presented in the previous section captures the various interfaces provided by the application's artefacts in order to support and enable such selections. The application based on context can adapt to the information provided to the user; for example if a user tries to hear a message sent by his child on the mobile phone in a noisy environment the application may adjust the volume.

Similarly the context can determine the most appropriate interface when a service is enabled. Imagine that a user is in a meeting and an SMS is received by his mobile phone; even though he may have forgotten to enable the phone's silent profile, the application can select to enable the vibration interface instead of the auditory one based on the context about place and activity. Another example is the following: a user is with his children in a museum and he receives a high priority e-mail and the display of his PDA is too small for him to read the whole document that a colleague sent him; the application tries to identify a larger display to present the document based on proximate artefacts' context and taking into account environmental parameters, like whether there are other users close to it, and issues of privacy and security, like whether the document is confidential.

This infrastructure could also be useful for people with special needs. Consider how useful a museum guide application could be if it can provide more auditory information or even a model that the user can touch when it identifies a user with impaired vision entering a gallery.

Another aspect of mobile applications is that they are used simultaneously by several users. The mobile application has to consider the number of users and their preferences and attempt to form groups of people with similar profiles and interests. The application can base its decisions on place context when many users exploit it. In a museum guide, it is easier to form groups of people with similar interests than in city guides. People in social places are more willing to share artefacts and services than in private spaces. In a home application the system can give priority to a father to print his last version of a work instead of first printing a child's painting, whereas in a work environment application it is arguable whether the boss should have greater priority.

An important issue in mobile applications is system failure because of device unavailability; a mobile phone may run out of battery or be out of range. The service classification represented in the proposed context-ontology can handle such situations, as it merely needs to identify another artefact that provides the same or similar services, therefore is abstracting the user from such problems.

Ubiquitous and mobile interfaces must be proactive in anticipating needs, while at the same time working as a spatial and contextual filter for information so that the user is not inundated with requests for attention (Brumitt, Meyers, Krumm et al., 2000). At the same time, ubiquitous interfaces must allow the user control over the interface (Abowd & Mynatt, 2000). Barkhuus and Dey (2003) presented an interesting case study on some hypothetical mobile phone services and have shown that users prefer proactive services to personalized ones. Providing proactive context aware services based on perceived user context is one of the major focuses of mobile and ubiquitous computing. However, proactive systems involving multiple smart artefacts often create complex problems if their behavior is not inline with user preferences and implicit understandings.

The ontology-based context model that we propose empowers users to compose their own personal mobile applications. In order to compose their applications they first have to select the artefacts that will participate and establish their associations. They set their own preferences by associating artefacts, denoting the sources of context that artefacts can exploit and defining the interpretation of this context trough rules in order to enable various services. As the context acquisition process is decoupled from the context management process, users are able to create their own mobile applications avoiding the problems emerging from the adaptation and customisation of applications like disorientation and system failures. A similar approach is presented in Zhang and Bruegge (2004).

Finally context can also assist designers to develop mobile applications and manage various interfaces and interaction techniques. Easiness is an important requirement for mobile applications; by using context according to our approach, designers are abstracted from the difficult task of context acquisition and have merely to define how context is exploited from various artefacts by defining simple rules. Our approach presents an infrastructure capable of handling, substituting and combining complex interfaces when necessary. The rules applied to the application's context and the reasoning process support the application's adaptation. The presented ontology-based context model is easily extended; new devices, new interfaces as well as novel interaction techniques can be exploited into a mobile application by simply defining their descriptions in the ontology.

FUTURE TRENDS

A crucial question that emerges is what the future of user interaction techniques and interfaces in mobile and ubicomp applications is. Aarts (2004) presented that the ultimate goal of user interaction in such applications is realizing "magic." Watching the movie Matilda (DeVito, 1996), a number of interaction techniques that designers try to integrate into mobile applications are presented as magic; eyes blinking can lead to opening or closing of the blinds, simple gesture movements may open or close the windows and pointing at specific devices switches them on and off.

However, can ubiquitous and mobile computing enable forms of magic? The answer is yes. As Scott (2005) mentions "by embedding computing, sensing and actuation into everyday objects and environments, it becomes feasible to provide new abilities to users, allowing them to exert levels of control and sensing in the physical world that were not previously possible." All superhuman or magic powers related to mobile applications are closely connected with context as defined in the previous sections. When users establish associations among artefacts define how artefacts should react on various context changes; a form of telekinesis is implemented as devices are ubiquitously controlled. Teleathesia can also be implemented using context; having associated their mobile phone with their house, users can be informed via their phone if someone is entering or leaving house by merging place's and family members' context. When a user drives back to home, this context information about the user's activity can be presented via a toy's display to his child who is playing waiting to go to the zoo; thus telepresence is enabled by context. Precognition and postcognition abilities can also be supported by exploiting context; from system experience and artefacts's knowledge important results from the past can be concluded, whereas precognition is also feasible if users have particularly incorporated information into the applications about future meetings, appointments, and so forth.

Magic is not applicable only to user interaction and interfaces in mobile applications. The artefacts that will be created may embody forms of magic. Consider the Weasley's clock in the Harry Potter book series (Rowling), it presents information about each member to the family based on their current activity and state. Context could enable the design and development of such artefacts.

Ontologies will play an important role in context representation for mobile applications as well as rule-based infrastructures and inference engines will be exploited for context reasoning in such applications. However a number of critical questions arise. For example, the location where ontologies are stored is still in dispute. Various infrastructures propose general ontologies centrally stored, whereas others prefer smaller and application-specific ontologies stored in distributed locations. Concerning the context-reasoning based on rule-based infrastructures, the issue that emerges is whether existing inference engines are suitable for mobile applications or need we turn our focus on different, more light-weight systems.

A research opportunity within the domain of this topic is how various interaction techniques and interfaces can be classified and represented into the ontology-based context model in order to provide a more effective selection of interaction techniques. During the previous years a number of markup languages were created in order to represent and describe interfaces; we believe that ontologies are the most suitable formal model for representing interfaces for mobile applications. Additionally, a formal model of interfaces described by an ontology may also assist the evaluation of interfaces used in mobile applications.

It is evident that the progress made in the last decade in the field of context-awareness in mobile systems is significant; however, certain critical issues remain open. Proactive mobile applications need to be certain for the context information based on which they decide their reaction in order to be trusted by the users; furthermore, mobile applications are usually multi-user so privacy and security are crucial.

CONCLUSION

The objective of this chapter was to present how context can support user interaction in mobile applications. Context-aware applications exploit location information in order to deliver location-aware services; when a user is identified by the system, personalized and adaptive services are provided. Whenever the user activity can be determined, the infrastructure provides the user with a proactive system that transforms his environment to a smart one; when the environmental parameters can be exploited along with the activity the system can best adapt the conditions or select the most suitable interaction method and interface. More advanced scenarios of proactive systems can even accommodate for the failures of particular system components.

However, users are still hesitant to adopt context-aware systems. The major reason for this is the fear that control may be taken away from them (Barkhuus & Dey 2003). Also, the gap between human expectations and the abilities of context-aware systems is sometimes big, especially when systems must handle ambiguous and uncertain scenarios or when the context on which decisions are based is imperfect.

The ontology-based context model that we presented in a previous section offers the benefits that were described above. Additionally, it allows users to setup their own context-aware applications and define the way that artefacts react to changes, giving them at the same time the sense of retaining control over the system. The context-management process assesses the state of an artefact in a two step process; the low-level context may contain impure information that is refined in order to produce the high-level context. In our system the user is able to dynamically update the rules that define the environment; so he is capable of foreseeing possible exceptions.

ACKNOWLEDGMENT

I would like to deeply thank the various people who, during the several months in which this endeavor lasted, provided me with useful and helpful assistance.

As part of the research described in this chapter carried out in the e-Gadgets and PLANTS projects. I would like to thank all my fellow researchers in these projects; especially thank Achilles Kameas, Christos Goumopoulos, Irene Mavrommati, and all my colleagues in the DAISy team of the Research Unit 3 of the Research Academic Computer Technology Institute for their encouragement and patience throughout the duration of these projects.

I would like to thank the anonymous reviewers, who read an early (and rather preliminary) proposal of this chapter and provided me with helpful feedback and invaluable insights, as well as Joanna Lumsden, the editor of this book, for her personal invitation to me to contribute to this book and her support.

I would like to commend the interest and great job done by Dimitris Dadiotis and Ourania Stathopoulou, who reviewed and proofed this chapter.

Most important, to Dimitris, who put up with lost weekends and odd working hours.

REFERENCES

Aarts, E. (2004). Keynote speak. *Adaptive Hypermedia Conference 2004*, Eindhoven, Netherlands.

Abowd, G., & Mynatt, E. (2000). Charting past, present, and future research in ubiquitous computing. *ACM Transactions on Computer-Human Interaction, 7*(1), 29-58.

Abowd, G., Bobick, A., Essa, I., Mynatt, E., & Rogers, W. (2002). The aware home: Developing technologies for successful aging. *Workshop held in conjunction with American Association of Artificial Intelligence (AAAI) Conference, Alberta, Canada.*

Ballagas, R., Borchers, J., Rohs, M., & Sheridan, J. G. (2006). The smart phone: A ubiquitous input device. *IEEE Pervasive Computing, 5(1),* 70-77.

Bardram, J., Hansen, T., Mogensen, M., & Soegaard, M. (2006). Experiences from real-world deployment of context-aware technologies in a hospital environment. *In proceedings of Ubicomp 2006* (pp. 369-386). Orange County, CA.

Barkhuus, L., & Dey, A. K. (2003). Is context-aware computing taking control away from the user? Three levels of interactivity examined. In *Proceedings of UbiComp 2003* (pp. 150-156). Springer.

Biegel, G., & Cahill, V. (2004, March 14-17). *A framework for developing mobile, context aware applications.* In 2nd IEEE Conference on Pervasive Computing and Communications. Orlando, FL

Bohnengerger, T., Jameson, A., Kruger, A., & Butz, A. (2002). User acceptance of a decision-theoretic location-aware shopping guide. In *Proceedings of the Intelligent User Interface 2002* (pp. 178-179). San Francisco: ACM Press

Brumitt, B., Meyers, B., Krumm, J., Kern, A., & Shafer, S. A. (2000). EasyLiving: Technologies for intelligent environments. *In proceedings of the 2nd international symposium on Handheld and Ubiquitous Computing* (pp.12-29). Bristol, UK.

Burrell, J., Gay, G. K., Kubo, K., & Farina, N. (2002). Context-aware computing: A test case. In *Proceedings of Ubicomp 2002* (pp. 1-15).

Chen, G., & Kotz, D. (2000). *A survey of context-aware mobile computing research*. (Tech. Rep. TR2000-381). Department of Computer Science, Dartmouth College.

Chen, H., Finin, T., & Joshi, A. (2004). An ontology for context aware pervasive computing environments. *Knowledge Engineering Review—Special Issue on Ontologies for Distributed Systems*. Cambridge: Cambridge University Press.

Christensen, J., Sussman, J, Levy, S., Bennett, W. E., Wolf, T. V., & Kellogg, W. A. (2006). Too much information. *ACM Queue, 4*(6).

Christopoulou, E., & Kameas, A. (2005). GAS Ontology: An ontology for collaboration among ubiquitous computing devices. *International Journal of Human-Computer Studies, 62*(5), 664-685.

Christopoulou, E., Goumopoulos, C., & Kameas, A. (2005). An ontology-based context management and reasoning process for UbiComp applications. *In proceedings of the 2005 joint conference on Smart objects and ambient intelligence: innovative context-aware services: usages and technologies* (pp. 265-270). Grenoble, France.

Crowley, J. L., Coutaz, J., Rey, G., & Reignier, P. (2002). Perceptual Components for Context Aware Computing. In *the proceedings of UbiComp 2002.*

Davies, N., Cheverst, K., Mitchell, K., & Efrat, A. (2001). Using and determining location in a context-sensitive tour guide. *In IEEE Computer, 34*(8), 35-41.

De Bruijn, J. (2003). *Using ontologies—Enabling knowledge sharing and reuse on the semantic Web.* (Tech. Rep. DERI-2003-10-29). Digital Enterprise Research Institute (DERI), Austria.

DeVito, D. (Director). (1996). Matilda. *Sony and TriStar Pictures.*

Dey, A. K. (2001). Understanding and using context. *Personal and Ubiquitous Computing, Special issue on Situated Interaction and Ubiquitous Computing, 5*(1), 4-7.

Dey, A.K., & Abowd, G.D. (2000). Towards a better understanding of context and context-awareness. *CHI 2000, Workshop on The What, Who, Where, When, Why and How of Context-awareness* (pp.1-6). ACM Press

Dey, A. K., Salber, D., & Abowd, G. D. (2001). A conceptual framework and a toolkit for supporting the rapid prototyping of context-aware applications. *Human-Computer Interaction Journal, 16*(2-4), 97-166.

Dix, A., Rodden, T., Davies, N., Trevor, J., Friday, A., & Palfreyman, K. (2000). Exploiting space and location as a design framework for interactive mobile systems. *ACM Transactions on Computer-Human Interaction, 7*(3), 285-321.

Erickson, T. (2002). Some problems with the notion of context-aware computing. *Communications of the ACM, 45*(2), 102-104.

Gellersen, H.W., Schmidt, A., & Beigl, M. (2002). Multi-sensor context-awareness in mobile devices and smart artefacts. *Mobile Networks and Applications, 7,* 341-351.

Goumopoulos, C., Christopoulou, E., Drossos, N. & Kameas, A. (2004). The PLANTS System: Enabling Mixed Societies of Communicating Plants and Artefacts. In *proceedings of the 2nd European Symposium on Ambient Intelligence* (pp. 184-195). Eindhoven, the Netherlands.

Gruber, T. G. (1993). A translation approach to portable ontologies. *Knowledge Acquisition, 5*(2), 199–220.

Häkkilä, J., & Mäntyjärvi, J. (2005). Collaboration in context-aware mobile phone applications. In *Proceedings of the 38th International Conference on System Sciences.* Hawaii

Henricksen, K., & Indulska, J. (2006). Developing context-aware pervasive computing applications: Models and approach. *Journal of Pervasive and Mobile Computing,* 2(1), 37-64.

Henricksen, K., Indulska, J., & Rakotonirainy, A. (2002). Modeling context information in pervasive computing systems. In F. Mattern & M. Naghshineh (Eds.), *Pervasive 2002* (pp. 167– 180). Berlin: Springer Verlag.

Hinkley, K., Pierce, J., Sinclair, M., & Horvitz, E. (2000). Sensing techniques for mobile interaction. *In CHI Letters, 2*(2), 91-100.

IST Advisory Group (ISTAG). (2001). *Scenarios for Ambient Intelligence in 2010-full.* http://www.cordis.lu/ist/istag-reports.htm

Jang, S., Ko, E. J., & Woo, W. (2005). Unified context representing user-centric context: Who, where, when, what, how and why. In *proceedings of International Workshop ubiPCMM05.* Tokyo, Japan.

Jones, M., & Marsden, G. (2005). *Mobile interaction design.* John Wiley & Sons.

Kindberg, T., Barton, J., Morgan, J., Becker, G., Caswell, D., Debaty, P., Gopal, G., Frid, M., Krishnan, V., Morris, H., Schettino, J., Serra, B., & Spasojevic M. (2002). People, places, things: Web presence for the real world. *Mobile Networks and Applications, 7*(5), 365–376.

Korpipää, P., Häkkilä, J., Kela, J., Ronkainen, S., & Känsälä, I. (2004). Utilising context ontology in mobile device application personalisation. In *proceedings of the 3rd international conference on Mobile and ubiquitous multimedia* (pp.133–140).

Koile, K., Tollmar, K., Demirdjian, D., Shrobe, H., & Darrell, T. (2003). Activity zones for context-aware computing. In *proceedings of UbiComp 2003 conference* (pp. 90-106). Seattle, WA.

Leong, L. H., Kobayashi, S., Koshizuka, N., & Sakamura, K. (2005). CASIS: A context-aware speech interface system. In *Proceedings of the 10th international conference on Intelligent user interfaces* (pp.231-238). San Diego, CA.

Marmasse, N., & Schmandt, C. (2000). Location-aware information delivering with comMotion. In *Proceedings of HUC 2000* (pp.157-171). Springer-Verlag.

Mitchell, K., Race, N. J.P., & Suggitt, M. (2006). iCapture: Facilitating spontaneous user-interaction with pervasive displays using smart sevices. In *PERMID workshop at the Pervasive 2006*. Dublin, Ireland.

Moran, T. P., & Dourish, P. (2001). Introduction to this special issue on Context-Aware Computing. *Human Computer Interaction 16*(2-4), 1-8.

Partridge, K., Begole, J., & Bellotti, V. (2005, September 11). Evaluation of contextual models. In *Proceedings of the First Internaltional Workshop on Personalized Context Modeling and Management for UbiComp Applications.* Tokyo, Japan.

Ranganathan, A., & Campbell, R. (2003). An infrastructure for context-awareness based on first order logic. *Personal and Ubiquitous Computing*, 7(6), 353–364.

Raptis, D., Tselios, N. & Avouris, N. (2005). Context-based design of mobile applications for museums: a survey of existing practices. In *Proceedings of the 7th international Conference on Human Computer interaction with Mobile Devices &Amp; Services. MobileHCI '05, 111* (pp. 153-160).. *ACM Press.*

Rowling, J. K. Harry Potter book series. *Bloomsbury Publishing Plc.*

Rukzio, E., Leichtenstern, K., Callaghan, V., Holleis, P., Schmidt, A., & Chin, J. (2006). An experimental comparison of physical mobile interaction techniques: Touching, pointing and scanning. In *proceedings of the 8th International Conference UbiComp 2006*, Orange County, CA.

Sadi, S. H., & Maes, P. (2005). xLink: Context management solution for commodity ubiquitous computing environments. In *proceedings of International Workshop ubiPCMM05.* Tokyo, Japan.

Schilit, B., Adams, N., & Want, R. (1994). Context-aware computing applications. In *proceedings of the IEEE Workshop on Mobile Computing Systems and Applications* (pp.85-90). Santa Cruz, CA.

Schilit, B., & Theimer, M. (1994). Disseminating active map information to mobile hosts. *IEEE Network*, *8*, 22-32.

Schmidt, A. (2000). Implicit human computer interaction through context. *Personal Technologies, 4*(2-3), 191-199.

Schmidt, A. (2002). *Ubiquitous computing—Computing in context*. Unpublished Ph.D. thesis, Department of Computer Science, Lancaster University, UK.

Scott, J. (2005). UbiComp: Becoming superhuman. *In the UbiPhysics 2005 workshop, Designing for physically integrated interaction*. Tokyo, Japan.

Stahl, C. (2006). Towards a notation for the modeling of user activities and interactions within intelligent environments. In *proceedings of the 3rd International Workshop on the Tangible Space Initiative (TSI 2006).* In Thomas Strang, Vinny Cahill, Aaron Quigley (Eds.), *Pervasive 2006 Workshop Proceedings* (pp. 441-452).

Strang, T., & Linnhoff-Popien, L. (2004). A context modeling survey. In *proceedings of the 1st International Workshop on Advanced Context Modelling, Reasoning And Management* (pp. 33-40). Nottingham, UK.

Strang, T., Linnhoff-Popien, L., & Frank, K. (2003). CoOL: A context ontology language to enable contextual interoperability. In *LNCS 2893 Proceedings of 4th IFIP WG 6.1 International Conference on Distributed Applications and Interoperable Systems* (pp. 236–247). Paris, France.

Truong, K. N., Abowd, G. D., & Brotherton, J. A. (2001). Who, what, when, where, how: Design issues of capture & access applications. In *proceedings of the International Conference: Ubiquitous Computing (UbiComp 2001)* (pp. 209-224). Atlanta, GA.

Uschold, M., & Gruninger, M. (1996). Ontologies: Principles, methods, and applications. *Knowledge Engineering Review, 11*(2), 93–155.

Wang, X. H., Gu, T., Zhang, D. Q., & Pung, H. K. (2004). Ontology based context modeling and reasoning using OWL. *Workshop on Context Modeling and Reasoning at IEEE International Conference on Pervasive Computing and Communication*. Orlando, FL.

Weiser, M. (1991). The computing for the 21st century. *Scientific American, 265*(3), 94-104.

Weiser, M. (1993). Some computer science issues in ubiquitous computing. *Communications of the ACM*, *36*(7), 75-84.

Zhang, T., & Bruegge, B. (2004, August). *Empowering the user to build smart home applications.* Second International Conference on Smart homes and health Telematics. Singapore.

KEY TERMS

Ambient Intelligence (AmI): Implies that technology will become invisible, embedded in our natural surroundings, present whenever we need it, enabled by simple and effortless interactions, accessed through multimodal interfaces, adaptive to users and context and proactively acting.

Context: Any information that can be used to characterize the situation of entities (i.e., whether a person, place or object) that are considered relevant to the interaction between a user and an application, including the user and the application themselves.

Context-Aware Application: An application based on an infrastructure that captures context and on a set of rules that govern how the application should respond to context changes.

Mobile Computing: The ability to use technology in remote or mobile (non static) environments. This technology is based on the use of battery powered, portable, and wireless computing and communication devices, like smart mobile phones, wearable computers and personal digital assistants (PDAs).

Ontology: A formal, explicit specification of a shared conceptualisation. A tool that can conceptualise a world view by capturing general knowledge and providing basic notions and concepts for basic terms and their interrelations.

Ubiquitous Computing (Ubicomp): Technology that is seamlessly integrated into the environment and aids human in their everyday activities. The embedding computation into the environment and everyday objects will enable people to interact with information-processing devices more naturally and casually than they currently do, and in whatever locations or circumstances they find themselves.

Chapter XIII
Context-Awareness and Mobile Devices

Anind K. Dey
Carnegie Mellon University, USA

Jonna Häkkilä
Nokia Research Center, Finland

ABSTRACT

Context-awareness is a maturing area within the field of ubiquitous computing. It is particularly relevant to the growing sub-field of mobile computing as a user's context changes more rapidly when a user is mobile, and interacts with more devices and people in a greater number of locations. In this chapter, we present a definition of context and context-awareness and describe its importance to human-computer interaction and mobile computing. We describe some of the difficulties in building context-aware applications and the solutions that have arisen to address these. Despite these solutions, users have difficulties in using and adopting mobile context-aware applications. We discuss these difficulties and present a set of eight design guidelines that can aid application designers in producing more usable and useful mobile context-aware applications.

INTRODUCTION

Over the past decade, there has been a widespread adoption of mobile phones and personal digital assistants (PDAs) all over the world. Economies of scale both for the devices and the supporting infrastructure have enabled billions of mobile devices to become affordable and accessible to large groups of users. Mobile computing is a fully realized phenomenon of everyday life and is the first computing platform that is truly ubiquitous. Technical enhancements in mobile computing, such as component miniaturization, enhanced computing power, and improvements in supporting infrastructure have enabled the creation of more versatile, powerful, and sophisticated mobile devices. Both industrial organizations and academic researchers, recognizing the powerful combination of a vast user population and a sophisticated computing platform, have focused tremendous effort on improving and enhancing the experience of using a mobile device.

Since its introducion in the mid-1980s, the sophistication of mobile devices in terms of the numbers and types of services they can provide has increased many times over. However, at the same time, the support for accepting input from users and presenting output to users has remained relatively impoverished. This has resulted in slow interaction, with elongated navigation paths and key press sequences to input information. The use of predictive typing allowed for more fluid interaction, but mobile devices were still limited to using information provided by the user and the device's service provider. Over the past few years, improvements to mobile devices and back-end infrastructure has allowed for additional information to be used as input to mobile devices and services. In particular, context, or information about the user, the user's environment and the device's context of use, can be leveraged to expand the level of input to mobile devices and support more efficient interaction with a mobile device. More and more, researchers are looking to make devices and services *context-aware*, or adaptable in response to a user's changing context.

In this chapter, we will define context-awareness and describe its importance to human-computer interaction and mobile devices. We will describe some of the difficulties that researchers have had in building context-aware applications and solutions that have arisen to address these. We will also discuss some of the difficulties users have in using context-aware applications and will present a set of design guidelines that indicate how mobile context-aware applications can be designed to address or avoid these difficulties.

What is Context-Awareness

The concept of context-aware computing was introduced in Mark Weiser's seminal paper 'The Computer for the 21st Century' (Weiser, 1991). He describes ubiquitous computing as a phenomenon '*that takes into account the natural human environment and allows the computers themselves to vanish into the background.'* He also shapes the fundamental concepts of context-aware computing, with computers that are able to capture and retrieve context-based information and offer seamless interaction to support the user's current tasks, and with each computer being able to *'adapt its behavior in significant ways'* to the captured context.

Schilit and Theimer (1994a) first introduce the term *context-aware computing* in 1994 and define it as software that "adapts according to its location of use, the collection of nearby people and objects, as well as changes to those objects over time." We prefer a more general definition of context and context-awareness:

Context is any information that can be used to characterize the situation of an entity. An entity is a person, place or object that is considered relevant to the interaction between a user and an application, including the user and applications themselves, and by extension, the environment the user and applications are embedded in. A system is context-aware if it uses context to provide relevant information and/or services to the user, where relevancy depends on the user's task. (Dey, 2001)

Context-aware features include using context to:

- Present information and services to a user
- Automatically execute a service for a user and
- Tag information to support later retrieval

In supporting these features, context-aware applications can utilize numerous different kinds of information sources. Often, this information comes from sensors, whether they are software sensors detecting information about the networked, or virtual, world, or hardware sensors detecting information about the physical world. Sensor data can be used to recognize the usage situation for instance from illumination, temperature, noise level, and device movements (Gellersen, Schmidt & Beigl, 2002; Mäntyjärvi & Seppänen, 2002). Typically, sensors are attached to a device and an application on the device locally performs the data analysis, context-recognition, and context-aware service.

Location is the most commonly used piece of context information, and several different location detection techniques have been utilized in context-awareness research. Global positioning system (GPS) is a commonly used technology when outdoors, utilized, for example, in car navigation systems. Network cellular ID can be used to determine location with mobile phones. Measuring the relative signal strengths of Bluetooth and WLAN hotspots and using the hotspots as beacons are frequently used techniques for outdoors and indoors positioning (Aalto, Göthlin,Korhonen et al., 2004; Burrell & Gay, 2002; Persson et al., 2003). Other methods used indoors include ultrasonic or infrared-based location detection (Abowd et al., 1997; Borriello et al., 2005).

Other commonly used forms of context are time of day, day of week, identity of the user, proximity to other devices and people, and actions of the user (Dey, Salber & Abowd, 2001; Osbakk & Rydgren, 2005). Context-aware device behavior may not rely purely on the physical environment. While sensors have been used to directly provide this physical context information, sensor data often needs to be interpreted to aid in the understanding of the user's goals. Information about a user's goals, preferences, and social context can be used for determining context-aware device behavior as well. Knowledge about a user's goals helps prioritize the device actions and select the most relevant information sources. A user's personal preferences can offer useful information for profiling or personalizing services or refining information retrieval. The user may also have preferences about quality of service issues such as cost-efficiency, data connection speed, and reliability, which relate closely to mobile connectivity issues dealing with handovers and alternative data transfer mediums. Finally, social context forms an important type of context as mobile devices are commonly used to support communication between two people and used in the presence of other people.

Relevance to HCI

When people speak and interact with each other, they naturally leverage their knowledge about the context around them to improve and streamline the interaction. But, when people interact with computers, the computing devices are usually quite ignorant of the user's context of use. As the use of context essentially expands the conversational bandwidth between the user and her application, context is extremely relevant to human-computer interaction (HCI). Context is useful for making interaction more efficient by not forcing users to explicitly enter information about their context. It is useful for improving interactions as context-aware applications and devices can offer more customized and more appropriate services than those that do not use context. While there have been no studies of context-aware applications to validate that they have this ability, anecdotally, it is clear that having more information about users, their environments, what they have done and what they want to do, is valuable to applications. This is true in network file systems that cache most recently used files to speed up later retrieval of those files, as well as in tour guides that provide additional information about a place of interest the user is next to.

Relevance to Mobile HCI

Context is particularly relevant in mobile computing. When users are mobile, their context of use changes much more rapidly than when they are stationary and tied to a desktop computing platform. For example, as people move, their location changes, the devices and people they interact with changes more frequently, and their goals and needs change. Mobility provides additional opportunities for leveraging context but also requires additional context to try and understand how the user's goals are changing. This places extra burden on the mobile computing platform, as it needs to sense potentially rapidly changing context, synthesize it and act upon it. In the next section, we will discuss the difficulties that application builders have had with building context-aware applications and solutions that have arisen to address these difficulties.

BUILDING MOBILE CONTEXT-AWARE APPLICATIONS

The first context-aware applications were centered on mobility. The Active Badge location system used infrared-based badges and sensors to determine the location of workers in an indoor location (Want et al., 1992). A receptionist could use this information to route a phone call to the location of the person being called, rather than forwarding the phone call to an empty office. Similarly, individuals could locate others to arrange impromptu meetings. Schilit, Adams and Want,(1994b) also use an infrared-based cellular network to location people and devices, the PARCTAB, and describe 4 different types of applications built with it (Schilit et al., 1994b). This includes:

- **Proximate selection:** Nearby objects like printers are emphasized to be easier to select than other similar objects that are further away from the user;
- **Contextual information and commands:** Information presented to a user or commands parameterized and executed for a user depend on the user's context;
- **Automatic contextual reconfiguration:** Software is automatically reconfigured to support a user's context; and
- **Context-triggered actions:** If-then rules are used to specify what actions to take based on a user's context.

Since these initial context-aware applications, a number of common mobile context-aware applications have been built: tour guides (Abowd et al., 1997; Cheverst et al. 2000; Cheverst, Mitchell & Davies, 2001), reminder systems (Dey & Abowd, 2000; Lamming & Flynn, 1994) and environmental controllers (Elrod et al., 1993; Mozer et al., 1995). Despite the number of people building (and re-building) these applications, the design and implementation of a new context-aware application required significant effort, as there was no reusable support for building context-aware applications. In particular, the problems that developers faced are:

- Context often comes from non-traditional devices that developers have little experience with, unlike the mouse and keyboard.
- Raw sensor data is often not directly useful to an application, so the data must be abstracted to turn it into useful context.
- Context comes from multiple distributed and heterogeneous sources, and this context often needs to be combined (or fused) to be useful. This process often results in uncertainty that needs to be handled by the application.
- Context is, by its very nature, dynamic, and changes to it must be detected in real time and applications must adjust to these constant changes in order to provide a positive user experience to users.

These problems resulted in developers building every new application from scratch, with little reuse of code or design ideas between applications.

Over the past five years or so, there has been a large number of research projects aimed at addressing these issues, most often trying to produce a reusable toolkit or infrastructure that makes the design of context-aware applications easier and more efficient. Our work, the Context Toolkit, used a number of abstractions to ease the building of applications. One abstraction, the context widget is similar to a graphical user interface widget in that it abstracts the source of an input and only deals with the information the source produces. For example, a location widget could receive input from someone manually entering information, a GPS device, or an infrared positioning system, but an application using a location widget does not have to deal with the details of the underlying sensing technology, only with the information the sensor produces: identity of the object being located, its location and the time when the object was located. Context interpreters support the interpretation, inference and fusion of context. Context aggregators collect all context-related to a specific location, object or person for easy access. With these three abstractions, along with a discovery system to locate and use the abstractions, an application developer no longer needs to deal with common difficulties in acquiring context and making it useful for an application, and instead can focus on how the particular application she is building can leverage the available context. Other similar architectures include JCAF (Bardram, 2005), SOCAM (Gu, Pung & Zhang, 2004), and CoBRA (Chen et al., 2004).

While these architectures make mobile context-aware applications easier to build, they do not address all problems. Outstanding problems needing support in generalized toolkits include representing and querying context using a common ontology, algorithms for fusing heterogeneous context together, dealing with uncertainty, and inference techniques for deriving higher level forms of context such as human intent. Despite these issues, these toolkits have supported and continue to support the development of a great number of context-aware applications. So, now that we can more easily build context-aware applications, we still need to address how to design and build *usable* mobile context-aware applications. We discuss this issue in the following section.

USABILITY OF MOBILE CONTEXT-AWARE APPLICATIONS

With context information being provided as implicit input to applications and with those applications using this context to infer human intent, there are greater usability concerns than with standard applications that are not context-aware. Bellotti and Edwards discuss the need for context-aware applications to be *intelligible*, where the inferences made and actions being taken are made available to end-users (Bellotti & Edwards, 2001). Without this intelligibility, users of context-aware applications would not be able to decide what actions or responses to take themselves (Dourish, 1997).

To ground our understanding of these abstract concerns, we studied the usability and usefulness of a variety of context-aware applications (Barkhuus & Dey, 2003a; 2003b). We described a number of real and hypothetical context-aware applications and asked subjects to provide daily reports on how they would have used each application each day, whether they thought the applications would be useful, and what reservations they had about using each application. All users were given the same set of applications, but users were split into three groups with each group being given applications with a different level of proactivity. One group was given applications that they would personalize to determine what the application should do for them. Another group was provided with information about how their context was changing, and the users themselves decided how to change the application behavior. The final group was evaluating applications that autonomously changed their behavior based on changing context. Additional information was also gathered from exit interviews conducted with subjects.

Users indicated that they would use and prefer applications that had higher degrees of proactivity. However, as the level of proactivity increased, users had increasing feelings that they were losing control. While these findings might seem contradictory, it should be considered that owning a mobile phone constitutes some lack of control as the user can be contacted anywhere and at anytime; the user may have less control but is willing to bear this cost in exchange for a more interactive and smoother everyday experience. Beyond this issue of control, users had other concerns with regards to the usability of context-aware applications. They were concerned by the lack of feedback, or intelligibility, that the applications provided. Particularly for the more proactive versions of applications, users were unclear how they would know that the application was performing some action for them, what action was being performed, and why this action was being performed. A third concern was privacy. Users were quite concerned that the context data that was being used on mobile platforms could be used by service providers and other entities to track their location and behaviors. A final concern that users had was related to them evaluating multiple context-aware applications. With potentially multiple applications vying for a user's attention, users had concerns about information overload. Particularly when mobile and focusing on some other task, it could be quite annoying to have multiple applications on the mobile device interrupting and requesting the user's attention simultaneously or even serially.

In the remainder of this chapter, we will discuss issues for designing context-aware applications that address usability concerns such as these.

Support for Interaction Design

Despite all of the active research in the field of context-aware computing, much work needs to be done to make context-awareness applications an integral part of everyday life. As context-awareness is still a very young field, it does not have established design practices that take into account its special characteristics. The development of applications has so far been done primarily in research groups that focus more on proof-of-concept and short-term use rather than deployable, long-term systems. For most of these applications, the interaction design has rarely been refined to a level that is required for usable and deployable applications. Particularly for applications aimed at consumers and the marketplace, robustness, reliability and usability must be treated more critically than they are currently, as these factors will have a significant impact on their success.

Currently, the lack of existing high-quality, commercial, and publicly available applications limits our ability to assess and refine the best practices in interaction design of context-aware mobile applications. As there is very little experience with real-life use of these applications, the ability of developers to compare and iterate on different design solutions is very restricted. As user groups for a particular application mostly do not exist yet, much of the current research is based on hypothesized or simulated systems rather that actualized use situations. Knowledge of what device features people fancy and which they just tolerate, and when application features become insignificant or annoying, are issues that are hard to anticipate without studies of long-term real-life usage.

As with any other novel technology, bringing it to the marketplace will bring new challenges. Bringing context-awareness to mobile devices as an additional feature may lead to situations where the interaction design is performed by people with little experience in context-aware computing. Using well-established commercial platforms such as mobile phones or PDAs often means that user interface designers only have experience with conventional mobile user interfaces. On the other hand, the technical specifications of an application are often provided by people who have no expertise in human-computer interaction issues. When entering a field that involves interdisciplinary elements, such as mobile context-awareness, providing tools and appropriate background information for designers helps them to recognize the risks and special requirements of the technology.

Hence, there are several factors which make examining context-awareness from the usability and interaction design perspective relevant. Failures in these may lead not only to unprofitable products, but may result in an overall negative effect—they may slow down or prevent the underlying technology from penetrating into mass markets.

Usability Risks for Mobile Context-Aware Applications

A system and its functionality are often described with mental models that people form from using the system. According to Norman (1990), one can distinguish between the designer's mental model and the user's mental model. The designer's model represents the designer's understanding and idea of the artefact being constructed, whereas the user's model is the user's conceptual model of the same artefact, its features and functionality, which has developed through her interaction with the system. In order to respond to the user's needs, efficiently fulfil the user's goals and satisfy the user's expectations, the designer's and user's understanding of the device or application should be consistent with each other, in other words, the user's model and designer's model should be the same (Norman, 1990).

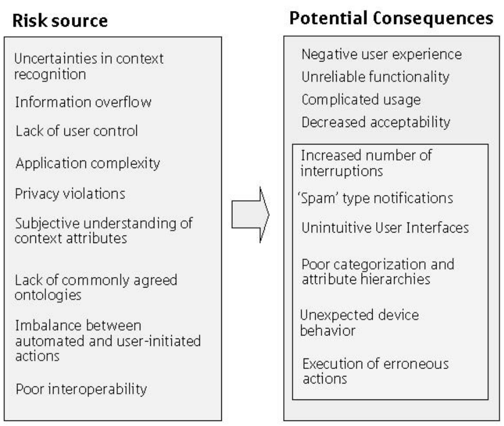
To ensure the best possible result, the mental models of different stakeholders in application development and use have to meet each other. First, the mental models of the application's technical designer and user interface designer should be consistent. This means that the user interface designer should have a basic understanding of the special characteristics of context-aware technology. Second, the designer's and user's mental models of the application should be the same. People's perception of context may differ significantly from each other, and both attributes and the measures used to describe context may vary greatly (Hiltunen, Häkkilä & Tuomela, 2005; Mäntyjärvi et al., 2003). The relationship between the designer's and the user's mental models should be checked with user tests several times during the design process. Without this careful design, there are two significant usability risks that may result: users will be unable to explain the behavior of the context-aware application, nor predict how the system will respond given some user action. While this is true of all interactive systems, it is especially important to consider for context-aware systems as the input to such systems is often implicit.

Context-awareness has several characteristics that can be problematic in interaction design. Figure 1 summarizes potential usability risks with context-aware applications.

A fundamental cause of potential usability risks is *uncertainty in context recognition*, which can be due to different reasons, such as detection accuracy, information fusion, or inferring logic. This is a key issue for designing the user interface for a mobile context-aware application, as it affects the selected features, their functionality and accuracy. In practice, features such as the proactivity level may be designed differently if the confidence level in context recognition can be estimated correctly. Uncertainty is a part of the nature of context-aware applications. Thus, it is important that the application and UI designers share a common understanding of the matter and take it into account when designing both the application and its user interface.

Application complexity has a tendency to grow when functions are added and it forms a potential risk for context-aware applications, as they use a greater number of information sources than traditional mobile applications. Hiding the complex nature of the technology while maintaining a sufficient level of feedback and transparency so that the user can still make sense of the actions the device is performing (i.e., intelligibility) is a challenging issue. Here, the involvement of user-

Figure 1. Sources of usability risks and their potential consequences related to context-aware mobile applications. Consequences that are unique to context-aware mobile applications are in the smaller rectangle on the right.



centric design principles is emphasized. Usability testing and user studies performed in an authentic environment combined with iterative design are key elements to producing well-performing user interface solutions.

Poor interoperability of services and applications relates to the absence of standardization in this maturing field and it limits the application design, available services, and seamless interaction desired across a wide selection of devices and users. Interoperability issues have gained much attention with the current trend of mobile convergence, where different mobile devices resemble each other more and more, yet providing services for them must be performed on a case-by-case basis.

Subjective understanding of context attributes creates a problem for user interface design, as the measures, such as the light intensity or noise level in everyday life are not commonly understood by end-users in terms of luxes or decibels but in relative terms such as 'dark,' 'bright,' 'silent' or 'loud.' This issue is connected to the *lack of commonly agreed ontologies*, which would guide the development of context-aware applications. The difficulties in categorizing context attributes and modeling context is evident from the literature (Hiltunen, Häkkilä & Tuomela, 2005; Mäntyjärvi et al., 2003).

As indicated earlier, *privacy violations* are possible with mobile context-aware systems collecting, sharing and using a tremendous amount of personal information about a user. When such information is shared with a number of different services, each of which will be contacting the user, *information overflow* often results. One can imagine a potential flow of incoming advertisements when entering a busy shopping street, if every shop within a radius of one hundred meters was to send an advertisement to the device. Information overflow is particularly a problem for the small screens that are typical with handheld devices.

As our earlier studies illustrated, the *lack of user control* can easily occur with mobile device automation, when context-triggered actions are executed proactively. However, the promise of context-awareness is that it provides "ease of use" by taking over actions that the user does not want to do or did not think to do for themselves. Any solution for correcting the *imbalance between the set of automated actions and user-initiated actions*, must take user control into account.

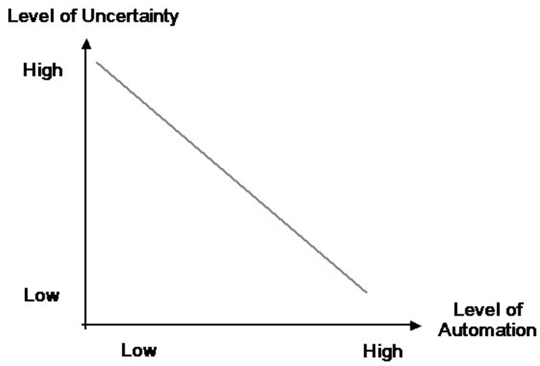
The consequences resulting from these usability risks are numerous. The general outcome can be a negative user experience. This may result from an increased number of interruptions, spam, and the execution of erroneous or otherwise unintuitive device behavior. Unreliable device functionality, and unintelligible user interfaces can lead to reduced acceptability of context-aware applications in the marketplace.

Design Guidelines for Mobile Context-Aware Applications

Context-awareness typically contains more risks than conventional, non-context-aware technology. At the same time, context-awareness can offer much added value to the user. In order to provide this value to end-users and avoid these negative design consequences and minimize usability risks, we have sought to provide a set of design guidelines that can offer practical help for designers who are involved in developing context-aware mobile applications (Häkkilä & Mäntyjärvi, 2006). These general guidelines have been validated in a series of user studies (Häkkilä & Mäntyjärvi, 2006) and should be taken into account when selecting the features of the application and during the overall design process.

GL1. Select appropriate level of automation. A fundamental factor with context-awareness is that it incorporates uncertainty. Uncertainty in context-recognition is caused by several different sources, such as detection accuracy, information fusion, or inferring logic. This is a key issue in designing user interfaces, as it affects the selected features, their functionality and accuracy. In practice, features such as the automation level or level of proactivity may be designed differently if the confidence level of context recognition can be estimated correctly. The relationship between uncertainty and selected application automation level is illustrated in Figure 2. As shown in Fig-

Figure 2. How uncertainty in context-recognition should affect the selected level of automation/proactivity



ure 1, *uncertainties in context recognition* create significant usability risks, however, by selecting an appropriate level of automation, an application designer can acknowledge this fact and address it appropriately. The greater the uncertainty is in the context-recognition, the more important it is not to automate actions. The automation level has also a direct relationship with user control, and its selection has a large impact on the number of expected interruptions the system creates for the user. The level of automation must be considered in relation to the overall application design, as it affects numerous issues in the user interface design.

GL 2. Ensure user control. The user has to maintain the feeling that he is in the control over the device. The user, who normally has full control over his mobile device, has voluntarily given some of it back to the device in order to increase the ease of use of the device. To address this *lack of user control*, an important usability risk, the user must be able to take control of the device and context-aware application at any time. The desire to take control can happen in two basic circumstances—either the device is performing erroneous actions and the user wants to take a correcting action, or the user just wishes to feel in control (a feeling that users often have). The user has to have enough knowledge of the context-aware application and the device functionality in order to recognize malfunctioning behavior, at least in the case where context-recognition errors lead to critical and potentially unexpected actions. The perception of user control is diminished if the device behaves in unexpected manner or if the user has a feeling that the device is performing actions without him knowing it. User control can be implemented, for example, with confirmation dialogues however, this must be balanced with the need to minimize unnecessary interruptions, our next guideline.

GL3. Avoid unnecessary interruptions. Every time the user is interrupted, she is distracted from the currently active task, impacting her performance and satisfaction with the system. In most cases, the interruption leads to negative consequences, however if the system thinks that the interruption will provide high value or benefit to the user, allowing the interruption is often seen as positive. Examples of this are reminders and alarm clocks. The user's interruptibility depends on her context and the user's threshold for putting up with intrusion varies with each individual and her situation. Some context-aware functionality is so important that the user may want the application to override all other ongoing tasks. This leads to a tension between avoiding unnecessary interruptions and supporting user control (GL2).

GL4. Avoid information overflow. The throughput of the information channel to each user is limited, and users can fully focus only on a small number of tasks at one time. In order to address the usability risk of *information overflow* where several different tasks or events compete for this channel, a priority ordering needs to be defined. Also, the threshold for determining the incoming event's relevancy in the context must be considered in order to avoid unnecessary interruptions (GL3). Systems should not present too much information at once, and should implement filtering techniques for to avoid messages that may appear to be spam to users. Also, information should be arranged in a meaningful manner to maintain and maximize the understandability of the system.

GL 5. Appropriate visibility level of system status. The visibility level of what the system is doing has to be sufficient for the user to be aware of the application's actions. While this guideline has been co-opted from Nielsen and Molich's user interface heuristics (1990), it has special meaning in context-aware computing. The implicit nature of context-awareness and natural *complexity of these types of applications* means that users may not be aware of changes in context, system reasoning or system action. When uncertainty in context-awareness is involved, there must be greater visibility of system state in order to allow the user to recognize the risk level and possible malfunctions. Important actions or changes in context should also be made visible and easily understandable for the user, despite the fact that users may have *subjective understandings of context attributes* and that there may be *no established ontology.* System status need not be overwhelming and interrupting to the user but can be provided in an ambient or peripheral fashion, where information is dynamically made more visible as the importance value grows, and may eventually lead to an interruption event to the user if its value is high enough.

GL 6. Personalization for individual needs. Context-awareness should allow a device or application to respond better to the individual user's personal needs. For instance, an application can implement filtering of interruptions according to the user's personal preferences. Personalization may also be used to improve the subjective understanding of context attributes. Allowing the user to name or change context attributes, such as location names or temperature limits, may contribute to better user satisfaction and ease of use. User preferences may change over time, and their representation in the application can be adjusted, for example implicitly with learning techniques or explicitly with user input settings.

GL 7. Secure user's privacy. *Privacy* is a central theme with personal devices, especially with devices focused on supporting personal communication, and impacts, for example trust, frequency of use, and application acceptability. Special care should be taken with applications that employ context sharing. Privacy requirements often vary between who is requesting the information, the perceived value of the information being requested and what information is being requested, so different levels of privacy should be supported. If necessary, users should have the ability to easily specify that they wish to remain anonymous with no context shared with other entities.

GL 8. Take into account the impact of social context. The social impact of a context-aware application taking an action must be part of the consideration in deciding whether to take the action or not. The application and its behavior reflects on users themselves. In some social contexts, certain device or user behavior may be considered awkward or even unacceptable. In such situations, there must be an appropriate *balance of user-initiated and system-initiated actions.* Social context has also has an effect on interruptibility. For example, an audible alert may be considered as inappropriate device behavior in some social contexts.

Once an application has been designed with these guidelines, the application must still be evaluated to ensure that the usability risks that have been identified for mobile context-aware systems have been addressed. This evaluation can take place in the lab, but is much more useful when conducted under real, *in situ*, conditions.

SUMMARY

Context-aware mobile applications, applications that can detect their users' situations and adapt their behavior in appropriate ways, are an important new form of mobile computing. Context-awareness has been used to overcome the deficit of the traditional problems of small screen sizes and limited input functionalities of mobile devices, to offer shortcuts to situationally-relevant device functions, and to provide location sensitive device actions and personalized mobile services.

Context-awareness as a research field has grown rapidly during recent years, concentrating on topics such as context-recognition, location-awareness, and novel application concepts. Several toolkits for enabling building context-aware research systems have been introduced. Despite their existence, there exist very few commercial or publicly available applications utilizing context-awareness. However, the multitude of research activities in mobile context-awareness allow us to make reasonable assumptions about tomorrow's potential applications. For example, navigation aids, tour guides, location-sensitive and context-sensitive notifications and reminders, automated annotation and sharing of photographs, use of metadata for file annotation, sharing or search are topics which frequently appear in the research literature and will likely be relevant in the future. In addition, using context-awareness to address the needs of special user groups, for example in the area of healthcare also appears to be a rich area to explore.

Despite the active research in context-awareness, there is much that remains to be addressed in interaction design and usability issues for context-aware mobile applications. Due the novelty of the field and lack of existing commercial applications, design practices for producing usable and useful user interfaces have not yet evolved, and end-users' experiences with the technology are not always positive. We have presented a set of 8 design guidelines which have been validated and evaluated in a series of user studies, which point to areas where user interface designers must focus efforts in order to address the usability issues that are commonly found with mobile context-aware applications.

While context-aware applications certainly have more usability risks than traditional mobile applications, the potential benefits they offer to end-users are great. It is important that application designers and user interface designers understand each other's perspectives and the unique opportunities and pitfalls that context-aware systems have to offer. With context-aware applications, careful application and interface design must be emphasized. The consequences resulting from usability risks include an overall negative user experience. Unsuccessful application design may result in diminished user control, increased number of interruptions, spam, and the execution of erroneous device actions or otherwise unintuitive behaviour. Unreliable device functionality and an unintuitive user interface can lead to decreased acceptability of the context-aware features in the marketplace.

In this chapter we have discussed the notion of context-awareness and its relevance to both mobile computing and interaction design in mobile computing. We have described technical issues involved in building context-aware applications and the toolkits that have been built to address these issues. Despite the existence of these toolkits in making context-aware applications easier to build, there are several additional issues that must be addressed in order to make mobile context-aware applications usable and acceptable to end-users. We have presented a number of design guidelines that can aid the designers of mobile context-aware applications in producing applications with both novel and useful functionality for these end-users.

REFERENCES

Aalto, L, Göthlin, N., Korhonen, J., & Ojala T. (2004). Bluetooth and WAP Push based location-aware mobile advertising system. In *Proceedings of the 2nd International Conference on Mobile Systems, Applications and Services* (pp. 49-58).

Abowd, G. D., Atkeson, C. G., Hong, J., Long, S., Kooper, R.. & Pinkerton, M. (1997). Cyberguide: a mobile context-aware tour guide. *ACM Wireless Networks, 3*, 421-433.

Bardram, J. (2005). The java context awareness framework (JCAF)—A service infrastructure and programming framework for context-aware applications. In *Proceedings of Pervasive 2005* (pp. 98-115).

Barkhuus, L., & Dey, A.K. (2003a). Is context-aware computing taking control away from the user? Three levels of interactivity examined. In *Proceedings of UBICOMP 2003* (pp. 149-156).

Barkhuus, L., & Dey, A.K. (2003b). Location-based services for mobile telephony: A study of users' privacy concerns. In *Proceedings of INTERACT 2003* (pp. 709-712).

Bellotti, V., & Edwards, K. (2001). Intelligibility and accountability: Human considerations in context-aware systems. *HCI Journal, 16*, 193-212.

Borriello, G., Liu, A., Offer, T., Palistrant, C., & Sharp, R. (2005). WALRUS: Wireless, Acoustic, Location with Room-Level Resolution using Ultrasound. In *Proceedings of the 3rd International Conference on Mobile systems, application and services (MobiSys'05)*, (pp. 191-203).

Burrell, J., & Gay, G. K. (2002). E-graffiti: Evaluating real-world use of a context-aware system. *Interacting with Computers, 14*, 301-312.

Chen, H., Finin, T. and Joshi, A. Chen, H., Finin, T., & Joshi, A. (2004). Semantic Web in the Context Broker Architecture. (2004). In *Proceedings of the Second IEEE international Conference on Pervasive Computing and Communications (Percom'04)*, (pp. 277-286).

Cheverst, K., Davies, N., Mitchell, K., & Friday, A. (2000). Experiences of developing and deploying a context-aware tourist guide: The GUIDE project. In *Proceedings of the 6th annual international conference on Mobile computing and networking (MobiCom)*, (pp. 20-31).

Cheverst, K., Mitchell, K., & Davies, N. (2001). Investigating Context-Aware Information Push vs. Information Pull to Tourists. In *Proceedings of MobileHCI'01*,

Dey, A.K., & Abowd, G.D. (2000). CybreMinder: A context-aware system for supporting reminders. In *Proceedings of the International Symposium on Handheld and Ubiquitous Computing* (pp. 172-186).

Dey, A.K., Salber, D., & Abowd, G.D. (2001). A conceptual framework and a toolkit for supporting the rapid prototyping of context-aware applications. *Human-Computer Interaction Journal 16*(2-4), (pp. 97-166).

Dourish, P. (1997). Accounting for system behaviour: Representation, reflection and resourceful action. In Kyng and Mathiassen (Eds.), *Computers and design in context* (pp. 145-170). Cambridge, MA: MIT Press.

Elrod, S., Hall, G., Costanza, R., Dixon, M., & des Rivieres, J. Responsive office environments. *Communications of the ACM* 36(7), 84-85.

Gellersen, H.W., Schmidt, A., & Beigl, M. (2002). Multi-sensor sontext-awareness in mobile devices and smart artefacts. *Mobile Networks and Applications, 7*, 341-351.

Gu, T., Pung, H.K., & Zhang, D.Q. (2004). A middleware for building context-aware mobile services. In *Proceedings of IEEE Vehicular Technology Conference* (pp. 2656-2660).

Häkkilä, J., & Mäntyjärvi, J. (2006). Developing design guidelines for context-aware mobile applications. In *Proceedings of the IEE International Conference on Mobile Technology, Applications and Systems.*

Hiltunen, K.-M., Häkkilä, J., & Tuomela, U. (2005). Subjective understanding of context attributes – a case study. In *Proceedings of Australasian Conference of Computer Human Interaction (OZCHI) 2005*, (pp. 1-4).

Lamming, M., & Flynn. M. (1994). Forget-me-note: Intimate computing in support of human memory. In *Proceedings of Friend21: International Symposium on Next Generation Human Interface* (pp. 125-128).

Mäntyjärvi, J., & Seppänen, T. (2002). Adapting applications in mobile terminals using fuzzy context information. In *Proceedings of Mobile HCI 2002* (pp. 95-107).

Mäntyjärvi, J., Tuomela, U., Känsälä, I., & Häkkilä, J. (2003). Context Studio—Tool for Personalizing Context-Aware Application in Mobile Terminals. In *Proceedings of Australasian Conference of Computer Human Interaction (OZCHI) 2003* (pp. 64-73).

Mozer, M.C., Dodier, R.H., Anderson, M., Vidmar, L., Cruickshank III, R.F., & Miller, D. The Neural Network House: An Overview. In L. Niklasson & M. Boden (Eds.), *Current trends in connectionism*, (pp. 371-380). Hillsdale, NJ: Erlbaum.

Nielsen, J., & Molich, R. (1990). Heuristic evaluation of user interfaces. In *Proceedings of CHI 1990* (pp. 249-256).

Norman, D. A. (1990). *The design of everyday things*. New York, NY: Doubleday.

Osbakk, P., & Rydgren, E. (2005). Ubiquitous computing for the public. In *Proceedings of Pervasive 2005 Workshop on Pervasive Mobile Interaction Devices (PERMID 2005)*, (pp. 56-59).

Persson, P., Espinoza, F., Fagerberg, P., Sandin, A., & Cöster, R. (2003). GeoNotes: A Location-based information System for Public Spaces. In K. Hook, D. Benyon & A. Munro (Eds.), *Readings in Social Navigation of Information Space* (pp. 151-173). London, UK: Springer-Verlag.

Schilit, B., & Theimer, M. (1994a). Disseminating active map information to mobile hosts. *IEEE Computer 8*(5), 22-32.

Schilit, B., Adams, N., & Want, R. (1994b). Context-aware computing applications. In *Proceedings of the IEEE Workshop on Mobile Computing Systems and Applications* (pp. 85-90).

Want, R., Hopper, A., Falcao, V., & Gibbons, J. (1992). The Active Badge Location System. *ACM Transactions on Information Systems*, *10*(1), 91-102.

Weiser, M. (1991). The computer for 21st century. *Scientific American,* 265(3), 94-104.

KEY TERMS

Context: Any information that can be used to characterize the situation of an entity. An entity is a person, place or object that is considered relevant to the interaction between a user and an application, including the user and applications themselves, and by extension, the environment the user and applications are embedded in.

Context-Awareness: A system is context-aware if it uses context to provide relevant information and/or services to the user, where relevancy depends on the user's task.

Design Guidelines: Guidelines or principles that, when followed, can improve the design and usability of a system.

Interaction Design: The design of the user interface and other mechanism that support the user's interaction with a system, including providing input and receiving output.

Mobile Context-Awareness: Context-awareness for systems or situations where the user and her devices are mobile. Mobility is particularly relevant for context-awareness as the user's context changes more rapidly when mobile.

Usability Risks: Risks that result from the use of a particular technology (in this case, context-awareness) that impact the usability of a system.

Chapter XIV
Designing and Evaluating In-Car User-Interfaces

Gary Burnett
University of Nottingham, UK

ABSTRACT

The introduction of computing and communications technologies within cars raises a range of novel human-computer interaction (HCI) issues. In particular, it is critical to understand how user-interfaces within cars can best be designed to account for the severe physical, perceptual and cognitive constraints placed on users by the driving context. This chapter introduces the driving situation and explains the range of computing systems being introduced within cars and their associated user-interfaces. The overall human-focused factors that designers must consider for this technology are raised. Furthermore, the range of methods (e.g., use of simulators, instrumented vehicles) available to designers of in-car user-interfaces are compared and contrasted. Specific guidance for one key system, vehicle navigation, is provided in a case study discussion. To conclude, overall trends in the development of in-car user-interfaces are discussed and the research challenges are raised.

INTRODUCTION

The motor car is an integral part of modern society. These self-propelled driver-guided vehicles transport millions of people every day for a multitude of different purposes, for example as part of work, for visiting friends and family, or for leisure activities. Likewise, computers are essential to many peoples' regular lives. It is only relatively recently that these two products have begun to merge, as computing-related technology is increasingly implemented within road-going vehicles. The functions of an in-car computing system can be broad, supporting tasks as diverse as navigation, lane keeping, collision avoidance, and parking. Ultimately, by implementing such systems car manufacturers aim to improve the safety, efficiency, and comfort and entertainment of the driving experience (Bishop, 2005)

Designing the user-interface for in-car computing systems raises many novel challenges, quite unlike those traditionally associated with interface design. For instance, in many situations, the use of an in-car system is secondary to the complex and already demanding primary task of safely controlling a vehicle in 2D space, whilst simultaneously maintaining an awareness of hazards, largely using the visual sense. Consequently, the level of workload (physical, visual, and mental) when using displays and controls becomes a critical safety-related factor. As a further example, in-car computing systems have to be used by a driver (and possible also, a passenger) who is sat in a constrained posture and is unlikely to be able to undertake a two handed operation. Therefore, the design (location, type, size, etc.) of input devices has to be carefully considered, accounting in particular for comfort, as well as safety, requirements.

This chapter aims primarily to provide the reader with an overall awareness of novel in-car computing systems and the key HCI design and evaluation issues. The focus is on the user-interface, that is, "the means by which the system reveals itself to the users and behaves in relation to the users' needs" (Hackos & Redish, 1998, p.5). Topics of relevance to both researchers and practitioners are raised throughout. Given the complexity of the driving task and the wide range of computing systems of relevance, the chapter principally provides breadth in its consideration of the subject. Nevertheless, some depth is explored in a case study investigation on the design and evaluation of user-interfaces for vehicle navigation systems.

TYPES OF IN-CAR COMPUTING SYSTEMS

Technology is increasingly being seen to have a critical role to play in alleviating the negative aspects of road transport, such as congestion, pollution and road traffic accidents (Bishop, 2005). Many technological initiatives are considered under the umbrella term, intelligent transport systems (ITS), where "ITS provides the intelligent link between travelers, vehicles, and infrastructure" (www.itsa.org, September, 2006). In this respect, in-vehicle computing systems are an important facet of ITS. Specifically, there are two core types of computing and communications systems which are either being implemented or developed for use in vehicles:

- **Information-based systems:** These systems provide information relevant to components of the driving environment, the vehicle or the driver. Examples of systems include navigation (facilitating route planning and following), travel and traffic information (traffic conditions, car parking availability, etc.), vision enhancement (providing an enhanced view of the road ahead, when driving at night, in fog or in heavy rain), driver alertness monitoring (informing the incapacitated driver if they are unfit to drive) and collision warnings (presenting warnings or advice regarding hazards).
- **Control-based systems:** These systems affect the routine, operational elements of the driving task. Examples of systems include adaptive cruise control (where the car is kept at a set time gap from a lead vehicle), speed limiting (the car speed cannot exceed the current limit), lane keeping (the driver's vehicle is kept within a given lane), self parking (vehicle automatically steers in low speed operation to position itself within a selected parking space) and collision avoidance (the vehicle automatically responds to an emergency situation). Clearly, such systems fundamentally change the nature of what we consider to be 'driving.'

It is important to note that there is a third category of in-car computing system, those which do not provide any functionality to support the driving task. These systems are an important consideration though, as they can negatively influence safety, particularly through the potential for distraction (Young, Regan & Hammer, 2003). Such systems may aim to enhance work-oriented productivity whilst driving (e.g., mobile phones, e-mail and Internet access) or be primarily conceived for entertainment and comfort purposes (e.g., music and DVD players, games). Moreover, they may be designed for dedicated use in a vehicle or for operation in a range of different contexts (often termed nomadic devices).

OVERALL HUMAN FACTORS ISSUES

Driving is a complex task involving a large number of subtasks that can be conceptualised as existing within three levels of an overall hierarchical structure (Michon, 1985):

- Strategic tasks (highest level global travel decisions—e.g., which car to take, which route to take);
- Tactical tasks (making concrete maneuvers requiring interaction with other road users—e.g., changing lane, turning at a roundabout);
- Operational tasks (motor execution of tasks planned at higher levels—e.g., turning steering wheel, pressing brake).

Inevitably, the introduction of new technologies into the driving context will have a considerable impact across all three levels. As a result, there are many human-focused issues that must be considered in the design and evaluation process for in-car computing systems. To provide structure to a discussion of these issues, two overall scenarios are envisaged which may arise from poor design and/or implementation of the technology.

- **Overload:** Many of these systems (particularly those providing novel types of information and/or interactions) lead to situations in which a driver must divide their attention between core driving tasks (e.g., watching out for hazards) and secondary system tasks (e.g., inputting information). Furthermore, systems may provide excessive information in an inappropriate way leading to high levels of mental workload, stress and frustration. Such issues often manifest themselves as distraction to the driver (biomechanical, visual, auditory and/or cognitive).
- **Underload:** Control-based systems clearly automate certain aspects of driving, transferring certain responsibilities from operator to computer (e.g., staying in lane), whilst potentially providing new tasks for the driver (e.g., monitoring system performance). Automation is a fundamental human factors topic with a considerable research literature (see Wickens et al., 2004). Key concerns in this context relate to the potential for a driver exhibiting reduced situational awareness (e.g., for other road users), negative behavioral adaptation (e.g., by taking greater risks) and de-skilling (e.g., driver not able to resume control in the event of system failure).

THE HUMAN-CENTRED DESIGN PROCESS

The fundamental components of a human-focused approach hold true for in-car computing, as much as for any interactive product or system, that is, early focus on users and tasks, empirical measurement and iterative design (Gould & Lewis, 1985). A comprehensive understanding of the context in which in-car computing devices will be used is especially important early in the design process. Context of use refers to “the users, tasks and equipment (hardware, software, and materials), and the physical and social environments in which a product is used” (Maguire, 2001, p.457). A context of use analysis assists in developing the initial requirements for a design and also provides an early basis for testing scenarios. Moreover, context of use analysis provides a focused approach that helps to ensure a shared view among a design team. In the driving situation, there are several context of use issues which will have a significant effect on how an in-car computing system is subsequently designed. Accounting for these raises many unique challenges for in-car user-interface designers.

Users

As with many other consumer products, there will be a large variability in user characteristics (e.g., in perceptual and cognitive abilities, computer experience, anthropometry) to consider when designing in-car computing systems. Car manufacturers may have particular socio-economic groups in mind when designing a vehicle, but the user base may still be extremely large.

One fundamental individual difference factor often addressed in research is driver age—drivers can be as young as 16 (in certain countries) and as old as 90. In this respect, younger drivers may be particularly skilled in the use of computing technology, in comparison with the population at large, but are especially prone to risk taking (Green, 2003). Moreover, studies have shown a limited ability to divide attention and prioritize sources of information, largely due to lack of driving experience (Wickman, Nieminem & Summala, 1998). Subsequently, system block outs, which prevent the use of complex functions in inappropriate driving situations, are likely to be of particular benefit for these individuals.

In contrast, older drivers often suffer from a range of visual impairments that can lead to a range of problems with in-vehicle displays. For instance, presbyopia (loss of elasticity in the lens of the eye) is extremely common amongst older people, as is reduced contrast sensitivity. Studies consistently show that older drivers can take 1.5 to 2 times longer to read information from an in-vehicle display compared to younger drivers (Green, 2003). Given that drivers have a limited ability to change the distance between themselves and an in-vehicle display, the size, luminance and contrast of presented information are obviously critical design factors.

Tasks

A key task-related issue is that the use of an in-car computing system is likely to be discretionary. Drivers do not necessarily have to use the system to achieve their goals and alternatives will be available (e.g., a paper map, using the brake themselves). As a result, the perceived utility of the device is critical. Furthermore, drivers' affective requirements may be particularly important. In certain cases, this requirement may conflict with safety-related needs, for instance, for a simple, rather than flashy or overly engaging user-interface.

The factor that most differentiates the driving context from traditional user-interface design is the multiple-task nature of system use, and in this respect, there are two critical issues that designers must take into consideration. The first concerns the relationship between primary driving tasks and secondary system tasks, as drivers seek to divide their attention between competing sources of information. Driving is largely a performance and time-critical visual-manual task with significant spatial components (e.g., estimating distances). Consequently, secondary tasks must not be overly time-consuming to achieve or require attentional resources that are largely visual, manual, and spatial in nature, if they are to avoid having a significant impact on primary driving.

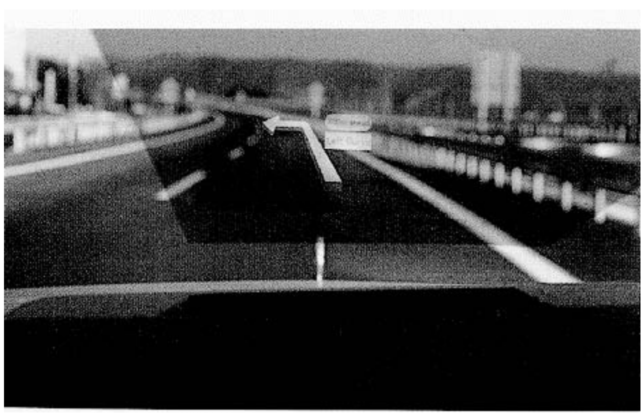
A second fundamental issue is the amount of information processing or decision making required for successful task performance, known as mental workload (Wickens et al., 2004). Novel in-car computing systems may provide functionality of utility to a driver or passengers, but interaction with the technology will inevitably increase (or in some cases decrease) overall workload. Context is very important here, as driving is a task in which workload varies considerably from one situation to another (compare driving in city traffic versus on the motorway). In this respect, certain authors (e.g., Green, 2004; Jones, 2002; Markkula, Kutila, & Engström, 2005) have taken the view that workload managers must be developed which make real-time predictions of the workload a driver is under and only present information or enable interactions to occur when overall workload is considered to be at an acceptable level. As an example, an incoming phone call may be sent straight to voice mail when the driver is considered to be particularly loaded (e.g., when driving in an unfamiliar city), but may be permitted in a lower workload scenario (e.g., driving along a dual carriageway and following a lead vehicle). Simple workload managers already exist in some vehicles (e.g., http://driving.timesonline.co.uk/article/0,,12929-2319048,00.html, September 2006), nevertheless, there are several complex research issues which must be addressed to fully realize the benefits of adaptive software in this context. For instance, workload managers need a comprehensive and accurate model of the driver, driving tasks and the driving environment. Given the vast range of variables of relevance to these categories, many of which do not lend to accurate and reliable measurement, extensive workload managers are likely to remain in the research domain for several years.

Equipment

The driving situation necessitates the use of input and output devices which are familiar to the majority of user-interface designers (pushbuttons, rockers, rotaries, LCDs, touchscreens, digitized or synthesized speech), together with equipment which is perhaps less known. For instance, there is a considerable research literature regarding the use of Head-Up Displays (HUDs) within vehicles. A HUD uses projection technology to provide virtual images which can be seen in the driver's line of sight through the front windscreen (see Figure 1). They are widely used within the aviation and military fields, and are now beginning to be implemented on a large-scale within road-based vehicles. HUDs will potentially allow drivers to continue attending to the road ahead whilst taking in secondary information more quickly (Ward & Parkes, 1994). As a consequence, they may be most applicable to situations in which the visual modality is highly loaded (e.g., urban driving), and for older drivers who experience difficulties in rapidly changing accommodation between near and far objects (Burns, 1999).

From a human-focused perspective, there are clear dangers in simply translating a technology from one context to another, given that vehicle-based HUDs will be used by people of varying perceptual and cognitive capabilities within an environment where there is a complex, continually changing visual scene. Specifically, researchers have established that poorly designed HUDs can mask critical road information, disrupt distance perception and visual scanning patterns, and negatively affect the ability of drivers to detect hazards in their peripheral vision (known as perceptual tunneling)—summarized by Tufano (1997) and Ward and Parkes (1994). Critical design factors that emerge from these findings include: display complexity; contrast and luminance; color choice; size of image; spatial location; and virtual image distance. Perhaps the most important design-related requirements are to consider carefully what and how much information is most appropriate to present on a HUD. There are temptations for designers to present ever-increasing amounts of information on HUDs. However, in contrast with traditional in-vehicle displays, a HUD image, by its very presence in the driver's line of sight, will demand focused attention (Burnett, 2003).

Figure 1. Example of a head-up display (HUD)



Environments

The physical environment is also a specific area that designers need to be aware of. In particular, the light, sound, thermal and vibration environment within a car can be highly variable. A range of design requirements will emerge from a consideration of these factors, for instance, potential for glare, problems with speech interfaces, use with gloves, and so on.

From anthropometric and biomechanical perspectives, the vehicle cabin environment provides many challenges for designers. This is an area in which designers make considerable use of CAD modeling to analyze different locations for displays and controls, ultimately aiming to ensure good fit for the design population. However, drivers sit in a constrained posture, often for several hours and have limited physical mobility (e.g., to comfortably view displays or reach controls). Consequently, there is limited space within a vehicle for the placement of a physical user-interface, a key problem for designers hoping to implement additional functionality within the vehicle.

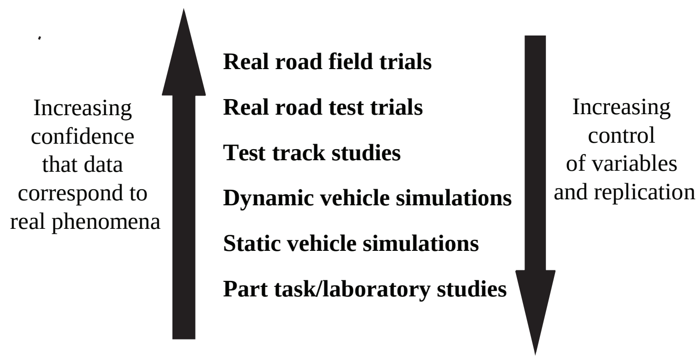
To a large extent, this factor has fueled the development of multi-modal user-interfaces, where a small number of controls, together with menu-driven screens, provide access to many functions within the vehicle. Clearly, such visually-oriented user-interfaces are likely to promote a considerable amount of "eyes-off-road" time, and empirical studies have confirmed this prediction (Dewar, 2002). Moreover, users mistaking the current mode is a well-established problem in user-interface design, and clear feedback is an important design requirement (Preece, Rogers & Sharp, 2002). In many respects, there is a trade-off in design between the number of discrete controls that a user must scan within a vehicle and the number of levels within a menu-based system that must be explored and understood. This is a very similar problem to that considered by HCI researchers in the 1980s and 1990s interested in the breadth versus depth of menus in graphical user-interfaces (Shneiderman, 1998). An overall recommendation from such HCI research is that breadth should generally be favored over depth, as excessive depth can cause considerably more problems for the user than an equivalent breadth, largely due to the cognitive problems of navigation (Shneiderman, 1998). Whilst such guidance is considered to be of relevance to the design of in-car computing, research is still required which considers the trade-off existing in the multiple-task driving environment.

METHODS FOR USE IN DESIGN AND EVALUATION

In considering the range of methods that a designer can utilize when designing and evaluating in-car computing systems, the first difficulty is in establishing what is meant by a method. In this respect, a "human factors method for testing in-car systems" can be seen to be a combination of three factors:

1. Which environment is the method used in (road, test track, simulator, laboratory, etc.). As can be seen in Figure 2 (redrawn and adapted from Parkes, 1991), there is a fundamental trade off in choosing a method environment between the need for control and the validity of results. Choosing an environment will also be largely influenced by practical considerations, the knowledge/skills of the design and evaluation team and resource limitations.
2. Which task manipulations occur (multiple task, single task loading, no tasks given, etc.)? In certain methods, there is an attempt to replicate or simulate the multiple task nature of driving. For other methods, performance and/or behavior on a single task may be assessed and the potential impact on other tasks inferred from this. Most removed from actual driving, some methods do not involve users, but instead aim to predict impacts or issues, for instance through the use of expert ratings or modeling techniques.
3. Which dependent variables (operationalized as metrics) are of interest. In assessing an in-car computing user-interface, a

Figure 2. Environments for evaluation of in-car computing devices and the relationship between validity and control



large range of possible metrics could be implemented. Some will relate to drivers' performance with primary driving tasks (e.g., lane position, hazard detection) or their use of primary vehicle controls (e.g., use of brake, steering wheel). Other metrics focus on driver performance and/or the demand of secondary tasks (e.g., task times, errors, display glances). As noted by Parkes (1991), usability evaluations of in-car computing devices should incorporate a wide range of measures relevant to the different levels of the driving task. For instance, at the strategic level, observation techniques and surveys are of relevance, whereas verbal protocols, interviews and questionnaires can capture the behavior of drivers at the tactical level. As noted by Parkes, such an approach provides "complete, rather than partial, pictures of product usability" (p.1445).

There is presently considerable research investigating specific methods for use in the design and evaluation of in-car user-interfaces. As noted by Preece et al. (2002), in deciding on any HCI method, the design team must consider the overall goals of the work, specific questions to be addressed, the practical and ethical issues and how data will need to be analyzed and reported. For in-car computing, these principles still hold, and many of the same global techniques used in the HCI area (for example, questionnaires, interviews, guidelines/checklists) will be used. However, by necessity, bespoke methods (or at least specific versions of generic methods) are required that account for the particular complex, safety-critical characteristics of the driving context. The following section summarizes key methods currently used and highlights some of the important research issues under investigation. Moreover, primary advantages and disadvantages are given. Table 1 summarizes some of the key issues.

Field Trials

Participants are given a car fitted with an operational system for several months for use in everyday activities. This method tends to look at broad issues relating to the long-term use of a system, for example, drivers' acceptance of the technology, and whether any behavioral adaptation effects arise. Objective data can be measured using on-board instrumentation (e.g., cameras, speed sensors) whereas subjective data is often captured using survey or interview-based approaches. Clearly, such a method provides an ecologically valid test of a system, and is particularly appropriate to the late stages of the design process where a robust prototype is available. Nevertheless, field trials can be extremely expensive and various ethical and liability considerations must be accounted for. An example of a field trial that was carried out in Sweden concerned drivers' use of intelligent speed adaptation systems (whereby a vehicle's speed is

Table 1. Overview of methods used to evaluate the user-interface for in-car computing systems

Method	Environment	Task manipulations	Overall Measures	Primary Advantages	Primary Disadvantages
Field trials	Real road (in everyday driving)	Multi-task (according to driver motivation)	Primary/ secondary task performance/ behavior, user opinions, etc.	Ecological validity, can assess behavioral adaptation	Resource intensive, ethical/liability issues to consider
Road trials	Real road (in pre-defined settings)	Multi-task (commonly, evaluator-manipulated)	Primary/ secondary task performance/ behavior, user opinions, etc.	Balance of ecological validity with control	Resource intensive, ethical/liability issues to consider
Simulator trials	Virtual driving environment (varying in fidelity)	Multi-task (commonly, evaluator-manipulated)	Primary/ secondary task performance/ behavior, user opinions, etc.	Control over variables, safe environment, cost-effective	Validity of driver behavior, simulator sickness
Occlusion	Laboratory/ statically in car	Secondary task achieved in controlled visual experience	Visual demand of user-interface	Standardized approach, control over variables	Limited scope, concern over validity of approach and metrics
Peripheral detection	Road/virtual driving environment	Multi-task (although commonly, evaluator-manipulated)	Visual/ cognitive workload	Assesses cognitive, as well as visual demand	Can be resource intensive, range of approaches
Lane change task	Specific lo-fidelity virtual driving environment	Multi-task motorway driving scenario	Primary lateral control of vehicle	Standardized approach, control over variables	Difficult to relate results to interface characteristics
15 second rule	Laboratory/ statically in car	Secondary task achieved without presence of driving task	Secondary task time (whilst stationary)	Simple approach	Only relates to certain aspects of visual demand
Keystroke-Level Model (KLM)	Modeling exercise	No user trials take place - models expert performance	Secondary task time (whilst stationary)	Quick/cheap, analysis explains results	Only relates to certain aspects of visual demand
Extended KLM	Modeling exercise	As for KLM, but with additional assumptions	Visual demand of user-interface	Quick/cheap, analysis explains results	Requires reliability assessments

automatically kept within the speed limit for the current area. Wallen, Warner & Aberg, 2005).

Road Trials

Drivers take part in a short-term (normally less than one day) focused study using a system in an instrumented car on public roads (occasionally on test tracks). For such trials, a wide range of variables may be measured and analyzed (e.g., visual behavior, workload, vehicle control, subjective preference) depending on the aims of the study. Road trials enable more experimental control than field trials, but are still potentially affected by a wide range of confounding variables (e.g., traffic conditions, weather). Furthermore, such a method remains costly to implement and requires robust protocols to ensure the safety of all concerned. Many road trials are reported in the literature, particularly concerning information and entertainment/productivity oriented systems. For instance, Burnett and Joyner (1997) describe a study which evaluated two different user-interfaces for vehicle navigation systems.

Simulator Trials

Drivers take part in a short-term (normally less than one day) focused study using a system fitted or mocked up within a driving simulator. The faithfulness that a simulator represents the driving task (known as its fidelity) can vary considerably, and configurations range from those with single computer screens and game controller configurations, through to real car cabins with multiple projections and motion systems. An example of a medium fidelity driving simulator is shown in Figure 3.

Driving simulators have become increasingly popular in recent years as a result of reduced hardware and software costs, and potentially offer an extremely cost-effective way of investigating many different design and evaluation issues in a safe and controlled environment (Reed & Green, 1999). Nevertheless, there are two key research issues concerning the use of driving simulators. Firstly, it is well known that individuals can experience symptoms of sickness in driving simulators, manifested as feelings of nausea, dizziness, and headaches. There has been considerable research regarding such sickness in virtual environments, and whilst there is still debate regarding the theoretical basis for the phenomenon (see for instance Nichols & Patel, 2002), there is practical guidance for those using driving simulators. For instance, screening questionnaires can be used to eliminate individuals who are most likely to experience sickness during a trial (Kennedy et al., 2001). Furthermore, various countermeasures can be used in the development of the simulator and its environment to reduce the prevalence of sickness (e.g., high, consistent frame rate, air-conditioning, natural background lighting. Nichols & Patel, 2002).

A second and more complex issue concerns validity, particularly behavioral (or construct) validity, that is, the extent to which drivers behave in the simulator as they would in the real world (Blaauw, 1982; Reed & Green, 1999). Driving simulator validity is problematic to study for several reasons. Running both road and simulator trials which are comparable (in terms of participants, tasks, measures, procedures, etc.) can be extremely difficult to achieve, and ultimately will be resource intensive. Furthermore, validity in this area is widely recognized to be a function of a large number of variables, including those relating to how the vehicle is represented (e.g., primary and secondary control design, the sense of enclosure, viewing angles, engine noise, vibration, motion, etc.) and those concerning the driving environment (e.g., visual field of view, screen resolution, graphical complexity, traffic representation, wind/road noise, etc. Kaptein et al.,1996; Peters & Peters, 2002). Most importantly, our understanding of validity must consider the driving task itself. Driving is a complex task, involving a substantial number of discrete physical, perceptual and cognitive behaviors, and a specific simulator configuration will only enable a subset of these to be investigated (e.g., speed control, headway maintenance).

As a consequence, despite the importance of the topic, there are few driving simulator validity studies in the open literature. Moreover, various limitations can be expressed for previous research in this area:

Figure 3. Example of a medium fidelity driving simulator

- It is difficult to generalize from existing validity studies, as they tend to be very specific to (a) the simulator configuration under investigation, and (b) the technology (hardware and software) available at that time (see for instance, Tornros (1998) compared with Blaauw (1982)).
- Studies inevitably only concern a small number of variables, for instance the effect of screen resolution and forward field of view on speed and headway choice (Jamson, 2001); or the effect of characteristics of torque feedback for steering on curve negotiation (Toffin et al., 2003).
- Studies often do not report critical data regarding the simulator configuration (e.g., steering sensitivity, max/min acceleration, driver eye height) which, for given types of study will be critical in results interpretation and cross-study comparison.

Occlusion

This is a laboratory-based method which focuses on the visual demand of in-vehicle systems. Participants carry out tasks with an in-vehicle system (stationary within a vehicle or vehicle mock up) whilst wearing computer-controlled goggles with LCDs as lenses which can open and shut in a precise manner (see Figure 4). Consequently, by stipulating a cycle of vision for a short period of time (e.g., 1.5 seconds), followed by an occlusion interval (e.g., 1.5 seconds), glancing behaviour is mimicked in a controlled fashion. Occlusion offers a relatively simple method of predicting visual demand, but is has been pointed out that its emphasis on user trials and performance data means that it requires a robust prototype and is therefore of limited use early in the design process (Pettitt et al., 2006).

Following considerable research, the occlusion method has recently been formalized as an international standard (ISO, 2005). In particular, guidance is given on how many participants are required, how much training to give, how many task variations to set, data analysis procedures, and so on. Moreover, two key metrics are stipulated: total shutter open time (the total time required to carry out tasks when vision is available); and resumability (the ratio of total shutter open time to task time when full vision is provided). For resumability, there is considerable debate regarding the merit of the measure. Advocates believe the metric provides an indication of the ease by which a task can be resumed following a period without vision (Baumann et al., 2004). Critics point out that the metric is also influenced by the degree to which participants are able to achieve tasks during occluded (non-vision) periods (Pettitt et al., 2006). Consequently, it can be difficult for a design team to interpret the results of an occlusion trial.

Peripheral Detection Task

This method requires drivers to carry out tasks with an in-car system (either on road or in a simulator)

Figure 4. The occlusion method with participant wearing occlusion goggles, with shutters open (left) and closed (right)

and to respond to the presence of lights within their periphery. The speed and accuracy of responses are considered to relate to the mental workload and distraction associated with secondary tasks (Young et al., 2003). The advantage of this method over occlusion is that it offers an assessment of cognitive, as well as visual demand (of relevance to the assessment of speech interfaces, for instance). The primary disadvantage is that the method still requires some form of driving task. Moreover, in contrast with occlusion, the method has not been fully standardized, and the ability to make cross study comparisons is severely limited by the specific choice of driving task scenarios (affecting task load and the conspicuity of the peripheral stimuli). It has also been noted that it is very difficult to discern between the level of cognitive demand and the visual demand for a given user-interface (Young et al., 2003).

An interesting recent development addresses some of these limitations. Engstrom, Aberg and Johansson (2005) considered the potential for the use of a haptic peripheral detection task, where drivers respond to vibro-tactile stimulation through the wrist whilst interacting with an in-vehicle system. Clearly, such a variation of peripheral detection is not affected by variations in lighting conditions. Furthermore, the authors argue on the basis of their validation work that this method provides "a 'pure' measure of cognitive load not mixed up with the effect of simply looking away" (p.233).

Lane Change Task

This method occurs in a basic PC simulated environment in which drivers are requested to make various lane change maneuvers whilst engaging with an in-vehicle system. The extent to which the profile of maneuver made by a driver varies from the optimum maneuver (the normative model) is considered to be a measure of the quality of their driving. Specifically, the method has the ability to assess the impact of an in-car computing system on a driver's awareness of the driving environment (perception, reaction), and, their ability to safely control the vehicle (maneuvering, lane keeping) Mattes (2003). Considerable research is ongoing with the lane change task in an attempt to develop an international standard (Transport Canada, 2006). Key research issues concern participant choice, training requirements and developing acceptable limits for performance.

15 Second Rule

Participants carry out tasks with an in-car computing system whilst stationary within a vehicle or mock up (i.e., with no driving task) and with full vision. The mean time to undertake a task is considered to be a basic measure of how demanding visually it is likely to be when driving (Green, 1999). A "cut-off" of 15 seconds has been set by the Society for Automotive Engineers (SAE). If the task on average takes longer than 15 seconds to achieve when stationary, it should not be allowed in a moving vehicle. The method is simple to implement and has the key advantage that it has been formalized in an SAE statement of best practice (SAE, 2000).

Research by Green (1999) and other research teams (e.g., Pettitt et al., 2006) has shown strong correlations between static task times and the total amount of time spent looking away from the road at displays/controls, both in simulator and road studies. However, the correlation between static task times and the duration of single glances towards an in-vehicle display is generally poor. This is important because a user-interface may promote a small number of very long glances (e.g., as a result of dynamically changing visual information) which can have a considerable negative effect on driving performance (Burnett & Joyner, 1997). It is for this primary reason that many authors advocate the use of the occlusion method as a better low-cost method for investigating the visual demand of an in-car user-interface (Pettitt et al., 2006; Stevens et al., 2004).

Keystroke Level Model (KLM)

The KLM method from the GOMs family of techniques is well known to HCI researchers and (to a lesser extent) practitioners (Preece et al., 2002;

Shneiderman, 1998). It is a form of task analysis in which system tasks with a given user-interface are broken down into their underlying physical and mental operators, e.g., pressing buttons, moving hand between controls, scanning for information. This is a method that is extremely cheap to implement, as there is no need for participants, and the method can be used with very basic prototypes early in the design process. Time values are associated with each operator and summed to give a prediction of task times. Researchers have developed new operator values relevant to the in-car situation (e.g., time to search a visual display, locate a control, move hand back to steering wheel) and have reported strong correlations between predicted task times and times based on user trials (Green, 2003; Pettitt et al., 2005). Task times can be related to certain measures of visual demand for in-car user-interfaces.

In an extension of the KLM method, Pettitt, Burnett and Stevens (2007) recently developed new rules that enable designers to develop predictions for a broader range of visual demand measures. In particular, the extended KLM considers a time-line view of an interaction in which a cycle of vision/non-vision occurs with a user-interface (similar to the occlusion protocol). The authors have found that their version of KLM can differentiate between tasks as effectively as does the occlusion technique, but recommend that further development is carried out to ensure that practitioners can utilize the method reliably.

CASE STUDY: VEHICLE NAVIGATION SYSTEMS

To ground many of the issues previously mentioned, a specific system type has been chosen for further discussion (vehicle navigation systems). Many of the individual points made for this system can be generalized and are applicable to other in-car computing technologies.

Vehicle navigation systems aim to support the strategic (e.g., route planning) and tactical (e.g., route following) components of the overall driving task. They have the greatest potential to assist drivers who undertake many unfamiliar journeys, for instance as part of work, or during leisure trips (e.g., when on holiday) and those who experience extreme difficulties with existing methods of navigation (particularly paper maps). When linked with reliable, real-time traffic information (thus providing dynamic guidance), the perceived utility of navigation systems to the everyday motorist is significantly enhanced (Bishop, 2005).

The market potential for vehicle navigation systems has already been demonstrated in Japan, where the technology has been available since the early 1990s. Approximately 40 percent of all vehicles on Japan's roads now have a navigation system installed (http://www.jetro.go.jp/en/market/trend/topic/2004_12_carnavi.html, September 2006). In many other countries, the popularity of navigation systems is currently reduced in relation to Japan, but is predicted to rise rapidly over the next few years (Bishop, 2005).

The majority of human factors issues relevant to this form of technology relate to overload, although as shall be seen, underload is increasingly being researched. With respect to overload, clearly, a key concern is the potential for driver distraction and there has been considerable research on this topic since the mid 1980s (see Young et al., 2003 and Srinivisan, 1999, for reviews). In using a vehicle navigation system, drivers must interact with controls (e.g., to enter a destination, chance map scale) and view/understand displays (e.g., to decide which turn to make, to examine options within a menu). In many cases, these interactions will arise when the vehicle is in motion. Consequently, to provide guidance for designers, researchers have aimed to understand how the user-interface design for a vehicle navigation system impacts on both navigating and primary driving performance. Specifically, research has aimed to answer the following three design-oriented questions:

What Information Should a Navigation System Provide?

To support route following, there are a wide range of different information types that a system could present, either referring to something real in the road environment (junction representations, street/road signs, landmarks, etc.) or indirectly referring to or pointing at aspects of the environment (distance to turn, directions, etc.). In this respect, researchers have established through a range of methodologies that the use of distinctive features of the environment (landmarks) within navigation instructions (e.g., "turn right at the church") offer considerable advantages over the use of distance to turn information (e.g., "turn right in 300 meters" Burnett, 2000; Ross, May & Grimsley, 2004). Moreover, research has identified the fundamental characteristics of landmarks which designers of vehicle navigation systems and providers of underlying map databases must consider in choosing appropriate landmarks for presentation by a navigation system (Burnett, Smith & May, 2001).

How Should Information be Presented?

Navigation and related information has to be presented to the driver in some way, and there has been considerable research on a range of topics. One key concern has been the impact of system modality (voice and/or visual) on driving and navigating performance. The general consensus here is the primary modality for presentation of navigation instructions should be auditory to reduce the conflict with the predominately visual driving task. However, information should also be presented visually, in particular, to support driver's understanding of more spatially complex maneuvers which cannot be represented easily in voice directions (Ross et al., 1995). Recently, Van Erp (2005) investigated empirically the potential for the use of passive touch as a novel modality for presentation of navigation instructions (specifically, vibro-tactile direction and distance to turn presented through the driver's seat). They concluded that haptic navigation displays offer various advantages over visual displays, for example, they provide a 'private' display to the driver appropriate for very simple maneuvers. Nevertheless, it must be noted that the authors did not make comparisons with the prevailing visual and auditory interfaces. Other research related to information presentation has considered a wide range of issues, such as the format of information (map-based vs. turn-by-turn based), the scheduling of information (when to present instructions), and the location of information (positioning of displays). On these topics, the reader is directed to Ross et al. (1995) and Srinivisan (1999).

How Should Drivers Interact with a Navigation System?

For drivers (or passengers) to interact with a vehicle navigation system, there must be a means by which they can enter data (e.g., postcode for an address), select from continuous/discrete options (e.g., voice volume levels, stored destinations), request/repeat information (e.g., voice directions), and move through the system (e.g., within and between menu screens). There is understandably a natural tendency for designers to utilise the familiar desktop computing paradigms, thus utilizing specific hardware devices (e.g., joysticks, touchscreens, buttons) and associated software approaches (e.g., use of menus, lists, scrolling). Historically, such paradigms were conceived as a means of overcoming the significant limitations of command-line user-interfaces and provided a what-you-see-is-what-you-get (WYSIWYG) experience for the user (Shneiderman, 1998). In the driving context, several studies have shown that such highly visual-manual user-interfaces can have a considerable impact on safety (Nowakowski, Utsui & Green, 2000; Tijerina, Palmer & Goodman, 1998).

As an alternative to such user-interfaces, speech shows promise as a largely non-visual/manual input method for navigation systems (Tsimhoni, Smith, & Green, 2002). Nevertheless, research has also shown that there is considerable potential for cognitive distraction with speech interfaces (Gärtner, König, & Wittig, 2001), and it is critical that recognition accuracy is very high. Moreover, designers must provide clear dialogue structures, familiar vocabulary, strong feedback and error recovery strategies. These issues are of particular importance given the potentially large number of terms (e.g., towns, street names) that might be uttered and the difficulties that a speech recognition system can experience with alphabet spelling (specifically, the 'e-set'—b, c, d, e, g etc.).

Recent research has also shown the potential for handwriting recognition in a driving context for inputting alphanumeric data (Burnett et al., 2005; Kamp et al., 2001). Whilst handwriting requires manual input, there is a reduced cognitive component and it is a more familiar method for users in contrast with speech interfaces. Nevertheless, issues relating to recognition accuracy remain and it is critical to place a handwriting touchpad in a location that facilitates the use of a driver's preferred hand (Burnett et al., 2005).

The difficulties for complex interactions with vehicle navigation systems are considered to be so significant that many authors believe that systems should disable "overly demanding" functionality when the vehicle is in motion (e.g., by "greying out" options when the vehicle is moving. Burnett, Summerskill & Porter, 2004; Green, 2003). This is currently a rich area for research, requiring an understanding of (a) what is meant by "overly demanding," (b) establishing valid/reliable metrics for the assessment of demand and finally, c) deciding where to put limits on acceptability (Burnett et al., 2004).

Underload for Vehicle Navigation Systems

In contrast with the overload perspective, over the last five years some researchers have viewed navigation systems as a form of automation, where underload issues become central. Vehicle navigation systems calculate a route for a driver according to pre-defined algorithms and then present filtered information, often via visual and auditory instructions. Two related concerns are emerging as important research questions, of particular relevance to user-interfaces which place a reliance on turn-by-turn guidance.

Firstly, it has been noted that there may be a poor calibration in the perceived versus objective reliability of in-car computing systems (Lee & See, 2004). This is of relevance as a navigation system (particularly the underlying digital map) that is unlikely ever to be 100 percent reliable. Nevertheless, drivers, largely based on their accumulated experience, may believe this to be the case. In certain situations, such overtrust in a system (commonly referred to as complacency) may lead to drivers following inappropriate routes and potentially making dangerous decisions, for instance, turning the wrong way down a one-way street. There is plenty of anecdotal evidence for such behavior in the popular press (e.g., http://www.timesonline.co.uk/article/0,,2-2142179,00.html, September 2006). Recently, research has replicated the effect in a simulated environment and indicated that there are considerable individual differences in the likelihood of a driver showing a complacency effect (Forbes & Burnett, 2007). Further research is considering what role the extended user-interface (training procedures, manuals, marketing information) can have in reducing complacency effects.

Secondly, drivers who use vehicle navigation systems may not develop a strong mental representation of the environments in which they travel, commonly referred to as a cognitive map. It has been stressed that traditional methods (e.g., using a paper map) require drivers to be active in the navigation task (route planning and following. Jackson, 1998; Burnett and Lee, 2005). Whilst the demands (particularly the cognitive demands) can initially be high, drivers who are engaged are able to develop landmark, then route knowledge, ultimately progressing to a map-like mental understanding (survey knowledge). Such a well-developed cognitive map means that drivers are able to navigate independent of any external source of information. Empirical research in this area has shown that drivers using current forms of user-interface for vehicle navigation system do indeed experience reduced environmental knowledge in relation to drivers using traditional methods (Burnett & Lee, 2005; Jackson, 1998). A key research question here is how user-interfaces can be developed which balance the need for low demands (workload) whilst simultaneously aiding drivers in developing a well formed cognitive map (Burnett & Lee, 2005).

FUTURE TRENDS AND CONCLUSION

The incessant growth in the use of cars and worries about road safety have led car manufacturers to offer more intelligent cars providing a range of novel functions to drivers. Moreover, existing mobile technologies such as PDAs, MP3 players, mobile phones, and so on, are increasingly being used within cars, as drivers seek to be more productive and to enjoy the time spent in their vehicles.

All of these computing-based systems offer potential benefits to drivers. This chapter has focused on some key design issues for user-interfaces from the perspective of the individual driver. However, as systems become commonplace within vehicles, there are fundamental conflicts to resolve between the requirements of an individual versus the overall traffic system. In this respect, the design of an in-car computing user-interface will be a critical consideration. As an example scenario, one can envisage many drivers using information systems providing the same information at the same time. Such a situation may lead to a range of problems, for instance the use of roads not designed for high volumes of traffic. Clearly, there is a need for overall management and an understanding of the impact that specific styles of user-interface will have on driver behavior.

A second broad issue for research concerns the interaction between multiple systems. This chapter has introduced the overload and underload concepts and discussed them in turn relating them to different individual systems. It is highly likely that in the short to medium term, overload will be given a prominent position in research and development work, whereas underload will emerge as an increasingly important topic in the medium to long term. However, this singular view neglects the fact that information and control-based systems are likely to be used together in a vehicle. Clearly, there will be various interaction effects for researchers to investigate. Moreover, there is a fundamental need to find the right balance between the two extremes of overload and underload. As noted by Dewar (2002, p. 330), “humans operate best at an optimal level of arousal, and either too much or too little workload can be detrimental to performance.”

The development of suitable methods for designing and evaluating in-car computing user-interfaces will continue to be an important research topic. Reliable and valid methods are required which are accepted within industry. A key motivation will be to establish ‘quick and dirty’ methods (and associated metrics) enabling designers to understand the likely demands of their user-interfaces early in the design process when very rudimentary prototypes are available. A further critical requirement is for “benchmarking,” that is, establishing a point of reference from which user-interfaces can be compared or assessed. Such benchmarks will be of particular benefit when identifying user-interface designs that are considered acceptable or unacceptable, particularly from a safety perspective.

REFERENCES

Baumann, M., Keinath, A., Krems, J.F., & Bengler, K (2004). Evaluation of in-vehicle HMI using occlusion techniques: Experimental results and practical implications. *Applied Ergonomics, 35*(3), 197-205.

Bishop, R. (2005). *Intelligent vehicle technology and trends.* London: Artech House Publishers.

Blaauw, G.J. (1982). Driving experience and task demands in simulator and instrumented car: A validation study. *Human Factors, 24*(4), 473-486.

Burnett, G.E. (2000). “Turn right at the traffic lights.” The requirement for landmarks in vehicle navigation systems. *The Journal of Navigation, 53*(3), 499-510.

Burnett, G.E. (2003). A road-based evaluation of a head-up display for presenting navigation information. In *Proceedings of HCI International conference (Human-Centred Computing)*, Vol. 3, (pp. 180-184).

Burnett, G.E., & Joyner, S.M. (1997). An assessment of moving map and symbol-based route guidance systems. In Y.I. Noy (Ed.), *Ergonomics and safety of intelligent driver interfaces* (pp. 115-137). Lawrence Erlbaum.

Burnett, G.E., & Lee, K. (2005). The effect of vehicle navigation systems on the formation of cognitive maps, In G. Underwood (Ed.), *Traffic and transport psychology: Theory and application* (pp. 407-418). Elsevier.

Burnett, G.E., Lomas, S., Mason, B., Porter, J.M., & Summerskill, S.J. (2005), Writing and driving: An assessment of handwriting recognition as a means of alphanumeric data entry in a driving context. *Advances in Transportation Studies, 2005 Special Issue*, 59-72.

Burnett, G.E., Smith, D., & May, A.J. (2001). Supporting the navigation task: Characteristics of 'good' landmarks. In Hanson, M.A. (Ed.), *Proceedings of the Annual Conference of the Ergonomics Society* (pp. 441-446). Taylor & Francis.

Burnett, G.E., Summerskill, S.J., & Porter, J.M. (2004). On-the-move destination entry for vehicle navigation systems: Unsafe by any means? *Behaviour and Information Technology, 23*(4), 265-272.

Burns, P. (1999). Navigation and the mobility of older drivers. *Journal of Gerontology. Series B, Psychological sciences and social sciences, 54*(1), 49-55.

Dewar, R.E. (2002). Vehicle design. In R.E. Dewar & P.L.Olson (Eds.), *Human factors in traffic safety,* (pp. 303-339). USA: Lawyers & Judges Publishing Company, Inc.

Engstrom, J., Aberg, N., & Johansson, E. (2005). Comparison between visual and tactile signal detection tasks applied to the safety assessment of in-vehicle information systems. In *Proceedings of the third international driving symposium on Human Factors in driver assessment, training and vehicle design*. (University of Iowa public policy center, Iowa).

Forbes, N., & Burnett, G.E. (2007). *Can we trust in-vehicle navigation systems too much?* In preparation, School of Computer Science and IT, University of Nottingham, UK.

Gärtner, U., König, W., & Wittig, T. (2001). Evaluation of manual vs. speech input when using a driver information system in real traffic. In *Proceedings of International driving symposium on human factors in driver assessment, training and vehicle design. (University of Iowa public policy center, Iowa)*, Available at: http://ppc.uiowa.edu/driving-assessment//2001/index.html

Gould, J. D., & Lewis, C. H. (1985). Designing for usability: key principles and what designers think. *Communications of the ACM, 28*(3), 300-311.

Green, P. (1999). The 15 second rule for driver information systems. In *Proceedings of the ITS America 9th annual meeting*, Washington, D.C.

Green, P. (2003). Motor vehicle driver interfaces. In J. A. Jacko & A. Sears (Eds.), *The human computer interactions handbook* (pp. 844-860). UK: Lawrence-Earlbaum Associates.

Green, P. (2004). Driver Distraction, telematics design, and workload managers: Safety issues and solutions. (SAE paper 2004-21-0022). *In Proceedings of the 2004 International Congress on Transportation Electronics (Convergence 2004, SAE publication P-387)*, (pp. 165-180). Warrendale, PA: Society of Automotive Engineers.

Hackos, J. T., & Redish, J. C. (1998). *User and task analysis for interface design*. NY: John Wiley and Sons.

ISO. (2005). *Road vehicles–Ergonomic aspects of transport information and control systems–Occlusion method to assess visual distraction due to the use of in-vehicle information and communication systems*. Draft International Standard ISO/DIS 16673. ISO/TC 22/SC 13.

Jackson, P. G. (1998). In search of better route guidance instructions. *Ergonomics, 41*(7), 1000-1013.

Jamson, H. (2001). Image characteristics and their effect on driving simulator validity. In *Proceedings of International driving symposium on Human Factors in driver assessment, training and vehicle design. (University of Iowa public policy center, Iowa).*

Jones, W.D. (2002). Building safer cars. *IEEE Spectrum*, *39*(1), 82-85.

Kamp, J.F., Marin-Lamellet, C., Forzy, J.F., & Causeur, D. (2001). HMI aspects of the usability of internet services with an in-car terminal on a driving simulator. *IATSS Research*, *2*5(2), 29-39.

Kaptein, N. A., Theeuwes, J., & Van der Horst, R. (1996). 'Driving simulator validity: Some considerations. *Transp. Res. Rec.*, *1550*, 30–36.

Kennedy, R. S., Lane, N. E., Stanney, K. M., Lanham, S., & Kingdon, K. (2001, May 8-10). Use of a motion experience questionnaire to predict simulator sickness. In *Proceedings of HCI International*, New Orelans, LA.

Lee, J.D., & See, K. (2004). Trust in automation: Designing for appropriate reliance. *Human Factors*, *46*(1), 50-80.

Maguire, M. (2001). Context of use within usability activities. *International Journal of Human-Computer Studies*, *55*, 453-483.

Markkula, G., Kutila, M., & Engström, J. (2005, November,14 - 16). *Online Detection of Driver Distraction - Preliminary Results from the AIDE Project.* Proceedings of International Truck & Bus Safety & Security Symposium Alexandria, VA.

Mattes, S. (2003). The lane change task as a tool for driver distraction evaluation. IHRA-ITS Workshop on Driving Simulator Scenarios, October 2003 - Dearborn, Michigan. Retrieved from www.nrd.nhtsa.dot.gov/IHRA/ITS/MATTES.pdf

Michon, J.A. (1985). A critical view of driver behaviour models. In L. Evans & R.S. Schwing (Eds.), *Human behaviour and traffic safety.* NY: Plenum Press.

Nichols, S., & Patel, H. (2002). Health and safety implications of virtual reality: A review of empirical evidence. *Applied Ergonomics*, *33*(3), 251-271.

Nowakowski, C., Utsui, Y., & Green, P. (2000). *Navigation system evaluation: The effects of driver workload and input devices on destination entry time and driving performance and their implications to the SAE recommended practice*. (Tech. Rep. UMTRI-2000-20). USA: The University of Michigan Transportation Research Institute.

Parkes, A.M. (1991). Data capture techniques for RTI usability evaluation. In *Advanced telematics in road transport—The DRIVE conference*: Vol. 2. (pp. 1440-1456). Amsterdam: Elsevier Science Publishers B.V.

Peters, G.A., & Peters, B.J. (2002). *Automotive vehicle safety.* London: Taylor and Francis.

Pettitt, M.A., & Burnett, G.E. (2005, November). Defining driver distraction. In *Proceedings of World Congress on Intelligent Transport Systems.* San Francisco.

Pettitt, M.A., Burnett, G.E., Bayer, S., & Stevens, A. (2006). Assessment of the occlusion technique as a means for evaluating the distraction potential of driver support systems. *IEE Proceedings in Intelligent Transport Systems*, *4*(1), 259-266.

Pettitt, M.A., Burnett, G.E., & Karbassioun, D. (2006). Applying the Keystroke Level Model in a driving context. In Bust, P.D. (Ed.), *Proceedings of the Ergonomics Society Annual conference, Taylor & Francis, Contemporary Ergonomics* 2006 (pp. 219-223).

Pettitt, M.A., Burnett, G.E., & Stevens, A. (2007). An extended Keystroke Level Model (KLM) for predicting the visual demand of in-vehicle information systems. In *Proceedings of the ACM Computer-Human Interaction (CHI) conference*. May 2007, San Jose, CA.

Preece, J., Rogers, Y., & Sharp. H. (2002). *Interaction design: Beyond human-computer interaction.* NY: John Wiley and Sons.

Reed, M.P., & Green, P.A. (1999). Comparison of driving performance on-road and in a low-cost simulator using a concurrent telephone dialling task. *Ergonomics, 42*(8). 1015-1037.

Ross, T., May, A.J., & Grimsley, P.J. (2004). Using traffic light information as navigation cues: implications for navigation system design. *Transportation research Part F, 7*(2), 119-134.

Ross, T., Vaughan, G., Engert, A., Peters, H., Burnett, G.E., & May, A.J. (1995). *Human factors guidelines for information presentation by route guidance and navigation systems (DRIVE II V2008 HARDIE, Deliverable 19).* Loughborough, UK: HUSAT Research Institute.

Shneiderman, B. (1998). *Designing the user-interface* (3rd edition). Addison Wesley.

Srinivisan, R. (1999). *Overview of some human factors design issues for in-vehicle navigation and route guidance systems.* Transportation Research Record 1694 (paper no. 99-0884). Washington, DC: National Academy Press.

Society of Automotive Engineers (SAE) (2000). *Navigation and Route Guidance Function Accessibility While Driving (SAE Recommended Practice J2364)*, version of January 20th, 2000. Warrendale, PA: Society of Automotive Engineers.

Stevens, A., Bygrave, S., Brook-Carter, N., & Luke, T. (2004). *Occlusion as a technique for measuring in-vehicle information system (IVIS) visual distraction: a research literature review.* Transport Research Laboratory (TRL) report no. TRL609. Crowthorne, Berkshire: TRL.

Tijerina, L., Palmer, E., & Goodman, M.J. (1998). *Driver workload assessment of route guidance system destination entry while driving.* (Tech. Rep. UMTRI-96-30). The University of Michigan Transportation Research Institute.

Toffin, D., Reymond, G., Kemeny, A., & Droulez, J. (2003, October 8-10). Influence of steering wheel torque feedback in a dynamic driving simulator. In *Proceedings of the Driving Simulation Conference North America*. Dearborn, USA.

Tornros, J. (1998). Driving behaviour in a real and a simulated road tunnel–A validation study. *Accident Analysis and Prevention, 30*(4), 497-503.

Transport Canada. (2006, June). *R Occlusion Research*. Presentation given to the AAM driver focus task force. Detroit.

Tsimhoni, O., Smith, D., & Green, P. (2002). *Destination entry while driving: Speech recognition versus a touch-screen keyboard*. (Tech. Rep. UMTRI-2001-24). The University of Michigan Transportation Research Institute.

Tufano, D.R. (1997). Automotive HUDs: The overlooked safety issues. *Human Factors, 39*(2), 303-311.

Van Erp, J.B.F. (2005). Presenting directions with a vibrotactile torso display. *Ergonomics, 48*(3), 302-313.

Wallen Warner, H.M., & Aberg, L. (2005). Why do drivers speed? In Underwood, G. (Ed.), *Traffic and transport psychology: theory and application* (pp. 505-511). Oxford, UK: Elsevier.

Ward, N. J., & Parkes, A.M.. (1994). Head-up displays and their automotive application: An overview of human factors issues affecting safety. *Accident Analysis and Prevention, 26*(6), 703-717.

Wickens, C.D., Lee, J.D., Liu, Y., & Becker, S.E.G., (2004). An introduction to human factors engineering (2nd Edition). New Jersey, USA: Pearson.

Wikman, A., Nieminen, T., & Summala, H. (1998). Driving experience and time-sharing during in-car tasks on roads of different width. *Ergonomics, 41*(3), 358-372

Young, K., Regan, M., & Hammer, M. (2003). *Driver distraction: a review of the literature.* Monash University Accident Research Centre, Report No. 206, November 2003.

KEY TERMS

Driver Distraction: Occurs when there is a delay by the driver in the recognition of information necessary to safely maintain the lateral and longitudinal control of the vehicle. Distraction may arise due to some event, activity, object or person, within or outside the vehicle that compels or tends to induce the driver's shifting attention away from fundamental driving tasks. Distraction may compromise the driver's auditory, biomechanical, cognitive or visual faculties, or combinations thereof (Pettitt & Burnett, 2005).

Driving Simulators: Provide a safe, controlled and cost-effective virtual environment in which research and training issues related to driving can be considered. Simulators vary considerably in their fidelity (i.e., the extent to which they replicate aspects of real driving).

In-Car Computing Systems: Provide information to support the driving task or control some aspect/s of the driving task. In-car computing systems may also provide information and/or services that are unrelated to driving.

Keystroke Level Model: Is an established HCI method used to predict expert's task times with a user-interface. It can be used with in-car user-interfaces to predict static task time, that is, the time taken to achieve tasks in a stationary vehicle. Recently, the KLM has been extended to predict visual demand measures related to the occlusion protocol.

Overload: (Due to in-car computing systems) occurs when a driver's information processing resources are overwhelmed and performance on primary driving tasks inevitably suffers.

Underload: (Due to in-car computing systems) occurs when automation of core driving tasks (such as steering, braking, etc.) has led to a situation in which driving performance has deteriorated. This may have arisen because the driver has reduced awareness of other road users, has changed their behavior in negative ways or has inferior skills/ knowledge in driving.

The Occlusion Protocol: Is a user trial method used in the design and evaluation of in-car user-interfaces. Participants typically wear LCD glasses which restrict the visual experience by only enabling short (e.g., 1.5 seconds) chunks of visual attention with an in-car user interface. Measures related to the visual demand of an interface can be established.

Chapter XV
Speech-Based UI Design for the Automobile

Bent Schmidt-Nielsen
Mitsubishi Electric Research Labs, USA

Bret Harsham
Mitsubishi Electric Research Labs, USA

Bhiksha Raj
Mitsubishi Electric Research Labs, USA

Clifton Forlines
Mitsubishi Electric Research Labs, USA

ABSTRACT

In this chapter we discuss a variety of topics relating to speech-based user interfaces for use in an automotive environment. We begin by presenting a number of design principles for the design of such interfaces, derived from several decades of combined experience in the development and evaluation of spoken user interfaces (UI) for automobiles, along with three case studies of current automotive navigation interfaces. Finally, we present a new model for speech-based user interfaces in automotive environments that recasts the goal of the UI from supporting the navigation among and selection from multiple states to that of selecting the desired command from a short list. We also present experimental evidence that UIs based on this approach can impose significantly lower cognitive load on a driver than conventional UIs.

INTRODUCTION AND BACKGROUND

The US census bureau reported in 2005 that the average American spends over 100 hours driving to and from work every year and spends several hundred more driving on errands, vacations, to social engagements, and so on. A significant fraction of this driving is spent while engaged in concurrent activities, such as listening to the radio, listening to music on a personal music player, operating an in-car navigation system, and talking on or accessing information with a hands-free or hand-held cell phone. These secondary activities involve interactions between the driver and a device that can distract the driver from the primary task—that of driving safely to the destination. While it is understood that the safest option is for a driver not to engage in such activities and instead concentrate completely on driving, drivers seem intent on engaging in these distractions; thus, minimizing the impact on safety is a worthy area of research.

It has been estimated that at least 25% of police reported accidents in 1995 involved some form of driver inattention (Wang, Knipling, & Goodman, 1996). A study by Stutts et al. (2001) estimated that, of the drivers whose state was known at the time of the crash, at least 13% were distracted, with adjusting the audio system of the car accounting for 11% of these distractions. Since the advent of cellular phone technology, there has been a great deal of research on the effects of cellular phone use on driving performance (e.g., Ranney et al., 2004); however, only recently have studies begun to address the effects of use of other in-car systems on driving performance. In an analysis of the 100-Car Naturalistic Driving Study, Klauer et al. (2006) found that "Drivers who are engaging in moderate secondary tasks are between 1.6 and 2.7 times as likely to be involved in a crash or near-crash, and drivers engaging in complex secondary tasks are between 1.7 and 5.5 times as likely" (p. 28).

Since these studies, a number of electronics manufacturers have introduced products that incorporate personal digital music collections into automobile audio systems. Some automobile manufacturers have gone as far as bundling a personal digital music player with the purchase of a new car. Recent high-end car models also offer GPS-linked navigation systems. These systems offer functions such as address entry and point-of-interest search, both of which are usually implemented as multi-step tasks requiring significant attention from the user. Navigation and entertainment systems are among the first examples of highly complex automotive interfaces that are available for use while driving. We expect the amount of information available in the car to continue increasing drastically as more and more car systems become networked, and as car makers try to differentiate their products by offering new functionality.

Given this situation, it becomes necessary to design effective user interfaces that will enable drivers to operate devices such as radios, music players, and cellphones in a manner that distracts them minimally from driving, while still allowing them to obtain the desired response from their devices.

A compelling choice for UI design in the automotive environment is the speech-based user interface. By "speech-based" we mean an interface which uses utterances spoken by the user as a primary input mode. A speech based interface may also have other input modes, such as dedicated or softkey input, and may also have voice feedback and/or visual feedback. By being largely hands free, a speech-based interface can minimize the need for the driver to disengage their hands from the steering wheel. By presenting information aurally, it can allow a driver to keep their eyes on the road.

These qualities are by themselves not sufficient: automobile UIs must not only allow drivers to keep their hands on the wheel and their eyes on the road, but also must allow them to keep their mind on the task at hand—that of driving safely. Spoken input is typically used as substitute for tactile input. It is frequently unclear how tactile actions such as turning a knob, pressing a button, or selecting an item on a touch screen may best be replaced by simple spoken commands that can be recalled easily by the user. The inability to recall the correct command can lead to poor system response and driver distraction. Further, Automatic Speech Recognition (ASR) engines are error-prone—they will often fail to recognize spoken input correctly, or worse still mistakenly recognize an incorrect command. These problems are magnified in noisy environments such as the inside of a fast-moving automobile.

Common approaches to minimize the adverse effects of ASR errors include detailed dialog mechanisms, help menus, confirmatory prompts, and error correcting dialogs. Unfortunately, these mechanisms are problematic in an automotive environment in which it is important for interactions between a driver and a system to be short in order to create minimal distraction to the driver.

Interface designers must strive to minimize unnecessary cognitive load such as those arising from extended interactions with a system, frustration from poor task completion, and other distractions of the mind. Recognition errors can often be minimized by constraining the choices that the ASR engine must consider at any one time. When using this approach, the UI must be designed to restrict the number of spoken commands available at each state while at the same time ensuring that the currently available commands are evident to a user. Thus, there is a strong bi-directional coupling between ASR performance and usability: ASR problems can manifest themselves to the end-user as usability problems, and interface design or implementation problems can easily lead to reduced recognition accuracy. Our experience has been that speech-based user interfaces are most effective when designed around the constraints of both ASR and UI. It is probably unrealistic to expect application and interface designers to be well versed in the technical details of ASR; however we believe that using a set of reasonable design guidelines could help automotive designers to design more effective speech-based interfaces.

There are a number of published documents that give design guidance for telematics interfaces, (see UMTRI, 2006); however, very few of these give any specific guidelines for the design of speech interfaces for telematics systems. Nor do they specify which general guidelines should apply to speech-based UIs, and which should not. In fact, some sets of guidelines (for instance, Society of Automotive Engineers, 2004) specifically exclude speech interfaces. Of course, there are also many sets of design guidelines for voice user interfaces in general, but few of these were written for the specific issues of an automotive user interface.

In this chapter, we will begin by presenting a number of design principles for the design of speech-based UIs, that have been derived from several decades of our combined experience in the development of spoken user interfaces for automobiles. Our discussion addresses issues such as the cognitive load imposed on the driver, interaction time, task completion rate, feedback, and the appeal of the system to the user. Based on these principles, we then present a brief review of the spoken UIs in a few current car models and highlight the positive and negative facets of these interfaces.

We conclude the chapter with a description of a speech interface paradigm which we call SILO (Divi et al., 2004) that conforms to most of our design principles, as an alternative to highly modal tree structured menus for the selection of a item out of a large number of alternatives. We describe an experiment that indicates that the SILO paradigm has significantly lower driving interference than the menu-based interface for a music selection task (Forlines et al., 2005).

DESIGN PRINCIPLES FOR SPEECH-BASED AUTOMOTIVE UIS

Here, we introduce a short set of design goals and specific recommendations for consumer automotive interfaces. These are based on the collective experience of the authors in implementing speech interfaces for use in automobiles over the last ten years. Most of the work that forms the basis for these recommendations is unpublished; however we feel that it is useful to present these recommendations here in collected form. Many of these guidelines can be found individually in other sources—this combined set is based on the particular constraints of consumer automotive applications.

Note that these guidelines are written with consumer automobile driving in mind. Commercial and military applications have different constraints and thus should be treated differently (for instance, the military user base receives training prior to in-field use, and the risk vs. task success rate balance may be different).

General Design Goals for Automotive Speech Interfaces

- **Reduce driver cognitive load.** In our view, the highest priority in design of automotive interfaces is to reduce the risk associated with performing any secondary task while driving. Since interfaces in use by a passenger can be distracting to the driver, the driver's cognitive load should be considered when designing interfaces that will be used by other occupants.
- **Reduce interaction time.** Reduced interaction time should lead to reduced risk.
- **Increase task completion rate.** Increased task completion rate should lead to reduced risk (fewer repeated interactions) and better user experience. Of course, task completion rate is related to the underlying speech recognition accuracy, but is also strongly influenced by the affordances offered by the user interface.
- The feedback (visual and audio) should reinforce correct use by the user.
- The user should be able to mentally model the system behavior.
- System should be effective for an experienced user (e.g., a long time owner).
- System should be appealing for a new user (especially during a pre-sales test drive).

These goals are given in order of decreasing importance. Not all of these design goals are achievable at the same time, and in fact, some may be in opposition to each other for various tasks. For instance, it can be difficult to make a system that is effective for both a new user and a long-time user. Below are some design recommendations that we believe follow from these design goals

Design Recommendations

1. **Interactions should be user paced:** Speech interfaces for use in automobiles should be *entirely* user paced. The primary task of the driver is the safe operation of the automobile. The operation of other equipment in the vehicle is a secondary task. The driver must be available to respond to changing traffic conditions. Driver responses to the system must be timed when the driver can spare attention for the secondary task. Thus every voice response by the driver should require its own push-to-talk event. Systems which ask for further voice input without waiting for a signal from the driver can cause the driver to feel pressured to respond even under difficult traffic conditions. Many state of the art systems include dialogs that violate this recommendation.
2. **Use a Push and Release button with a listening tone:** After the user activates the Push-to-talk, the system should promptly produce a short pleasant listening tone to indicate that it is listening. Studies have indicated that in the absence of a listening tone, some users will start speaking prior to activating the push-to-talk and other users will delay speaking for a variable amount of time. Either of these changes in timing can confuse the speech recognizer, resulting in an overall reduction of recognition accuracy.

   In the presence of a listening tone, most users quickly learn to wait for the listening tone and to start speaking promptly after they hear the tone. However, the time between activation of the push-to-talk and production of the listening tone must be short and consistent; otherwise the cognitive load on the user is higher.

   Push and Release interfaces, where the user releases the PTT button immediately, usually impose a lower cognitive load on the user than Push and Hold style interfaces, where the user is required to hold the button for the entire utterance. This is because the user's physical actions are sequential (press, then speak) in a Push and Release interface rather than simultaneous. It can be useful to indicate with a different tone that the system is no longer listening. This affordance helps the user to adapt to system constraints such as listening timeouts.

3. **Use physical input instead of voice for simple things:** There are numerous actions for which there are existing effective physical interfaces. In almost all instances keeping those physical interfaces is superior to the substitution of voice commands. For example, using up and down buttons to navigate through a list of options is much easier than saying commands such as "scroll down". In addition, the users are already familiar with such physical interfaces. There is also less new learning involved to be able to use the interface.
4. **Provide always-active commands including voice help:** A "Help" or "What can I say" command should always be available for users who are unsure as to what functions are currently available. Systems that have modal behavior should have mode-specific voice help, as well as commands to cancel the current mode and to backup.
5. **Use consistent grammars with minimal modality:** Consistency and predictability of the grammar is very important so that the user will have less to remember. Also the grammar should have minimal modality so that users will have access to the functions of the system with fewer interactions and will not need to remember the state of the system. Modal behavior should be associated with audio and visual cues so that the user can easily understand what mode the system is in.
6. **Visual cues should be consistent with the active grammar:** The use of visual cues that do not model the grammar usually causes an increased number of out-of-grammar utterances.
7. **Feedback should indicate the recognition result:** The visual and audio feedback given to the user should indicate what was heard by the speech recognizer. This helps to reduce confusion when the system does not behave as expected, either due to misrecognition, or to user confusion about the effect of the spoken command. Users have a strong tendency to mimic the sentences that they hear; thus, some systems echo the recognized utterance back to the user as a confirmation of what the system heard or echo the preferred form of the command to help the user learn the grammar.
8. **Reasonable behavior for out-of-grammar utterances:** Systems should attempt to detect out-of-grammar utterances and indicate that the last utterance was not understood. It is far better for the system to respond that the command was not understood than it is to perform an unexpected action. Unexpected actions cause user confusion (e.g., was the utterance in the grammar and misrecognized, or not in the grammar?). Lack of rejection raises the cost of an out-of-grammar utterance as the user must take whatever action is necessary to undo the undesired action, resulting in significantly longer task completion times.
9. **Provide reasonable backoff strategies:** In some cases a speech interface will consistently fail to recognize certain voice commands from the user. For example, a system may allow the entry of a street name as part of an address, a very difficult voice recognition problem for a locality with many streets. Thus, if the voice system fails to get the correct street after a couple of tries, it should offer an alternative method of entry, such a spelling. As another example, many systems allow the entry of long telephone numbers in one long utterance by the user. For some users the error rate for strings of ten or more digits can be too high. For such users the system should allow the entry and correction of the digits forming a long telephone number in smaller chunks.

EVALUATION OF RECENT AUTOMOTIVE SPOKEN USER INTERFACES

In April 2006, we conducted a review of three automotive speech-based interfaces from model year 2006. Rather than give an exhaustive review, here we highlight the system attributes that had the most impact on interface usability. In order to focus on the attributes of the interfaces rather than the identities of the manufacturers, we will call these interfaces "Model A," "Model B," and "Model C."

The state of the art for speech-based navigation systems at the time of the review was:

- Push and Release interface with listening tone
- Grammar based systems with varying degrees of modality
- Navigation Entry of part or all of an address by voice

There was significant variation in these features between the tested systems. There was also significant variation in the accuracy of the underlying automatic speech recognition engines, but we found the usability to be more affected by the UI design than the underlying ASR.

Model A

Model A had good ASR engine accuracy. However, there was no rejection of out-of-grammar utterances, so ASR errors (misrecognitions) frequently led to unexpected actions.

User Pacing

This system was entirely user-paced, with a push-to-talk button which was required for every interaction, and a prompt, pleasant listening tone.

Grammar Consistency and Modality

Model A had a fairly simple modal grammar structure, with some common always-active commands. The grammar format was consistent, and the audio feedback modeled the grammar. Each mode was associated with a visual state. Overall it was easy to understand what mode the system was in, and to guess what voice commands were available.

However, there were some modes with inconsistent design. Two interesting examples:

- **Voice help modes:** Each voice mode (v_n) had a corresponding help mode (vh_n) which was invoked through the always-active help command. Each help mode visually displayed the available commands for its parent mode in a numbered list.
  However, the help modes had their own very limited grammar. Thus in any help mode (vh_n) the commands for v_n were displayed but not active. The result of this was to induce users to utter unavailable commands in help mode, resulting in frequent and confusing misrecognition in help mode. In order to utter one of the displayed commands, users had to first dismiss the help screen, at which point the commands were no longer visible. This design was almost consistent with design recommendation 6 (*Visual cues should be consistent with the active grammar*), but the seeming minor detail, that the visual cues for these modes matched an inactive grammar, led to a major usability issue.
- **Setup modes:** This system had several preference modes for manipulating system settings. Typically, the system displayed a set of buttons and sliders when in one of these modes. There were voice commands for manipulating the values, but the commands were inconsistent and there was no visual indication of which objects could be manipulated by voice or what command language to use. There were physical/softkey methods

to change the settings, and it appeared that the speech interface had been grafted onto these modes as an afterthought. The interface could have been more effective without the additional voice commands, consistent with design recommendation 3 (*Use physical input instead of voice for simple things*).

Other Usability Issues

The interface included a 'Back' button, which was probably designed to move the interface to the previous state after a recognition event, but had inconsistent behavior. This was an interesting attempt to provide a reasonable backoff strategy for misrecognition. Prior speech user interface (SUI) implementations have shown that a 'Back' button or an "Undo that" command that undoes the previous action can be a very effective affordance for an ASR system, mitigating the inevitable ASR errors. In this particular case, the behavior was unreliable.

Model B

Model B had poor ASR engine accuracy, especially in high noise conditions typical of highway driving.

There was no rejection of out-of-grammar utterances, however, some modes accepted unavailable commands and then reported that that command was not available in the current mode.

User Pacing

This interface had a push-to-talk button which was required for every interaction. As noted in design recommendation 2 (*Use a Push and Release button with a listening tone*), this interface pattern usually improves usability, but not in this case.

Model B's implementation of push-to-talk degraded not only the usability of the interface, but also the ASR performance, all while increasing cognitive load: Activation of the push-to-talk was followed by a visual cue which indicated that the user could begin speaking. This visual cue was followed by a highly variable delay and then a listening tone. Sometimes the listening tone was produced promptly, other times there was a delay of 1-2 seconds after the visual cue. Our testing indicated that the variable listening tone delay frequently caused recognition errors (when the users spoke too quickly) and higher cognitive load (the user had to actively think about waiting for the tone). Furthermore, the listening tone was hard to hear in high noise conditions.

Grammar Consistency and Modality

Model B had a fairly simple modal structure. However, the voice modes were not well associated with visual cues. This made it very difficult to determine which commands were available at any given time. For instance, there were several modes where a map was visible on the screen, but only one of these was "map mode." Therefore the system sometimes responded "That command is only available in map mode" even when there was a map showing on the screen.

Model B also had a confusing lack of consistency between visual cues, voice commands, and voice feedback. Recognition feedback was aligned with, but did not match the grammar. When the user spoke a command of the form "action object," the voice response was "<action>-ing object." For instance:

- **User:** *"raise temperature"*
- **System:** *"raising temperature"*

This was a clear indication of what action the system was taking, but an indirect indication of what command had been recognized. The use of different language sometimes induced users to copy the response language (e.g., to say "raising temperature" by accident).

In those cases where users mimicked the response language, the lack of rejection of out-of-grammar utterances usually caused an unexpected action to occur, forcing the user to attempt to undo the unexpected action before re-trying the original task.

Model C

Model C had reasonably good ASR performance. It also had rejection of out-of-grammar utterances. Some out-of-grammar utterances were misrecognized, but many were rejected.

User Pacing

Model C was not user paced. An interaction with the system could consist of a single (conversational) turn by the user or an entire sequence of turns by user and system. Push-to-talk was used to initiate a task, but then the system controlled the turn-taking and pace until the end of the task. To mark the user's subsequent turns, the system generated a voice prompt followed by another listening tone and immediately expected more voice input from the user. This design was confusing to new users because they did not know when a conversational sequence would be finished.

When used during driving conditions, the interface created a high cognitive load on the user, who had to pay enough attention to the voice prompts to be able to respond to demands for more input. If the user ignored a prompt, the prompt and listening tone were repeated, thus during a task the system constantly demanded attention from the user regardless of the traffic conditions. Our experience indicates that this can be hazardous—see design recommendation 1 (*Interactions should be user paced*).

Grammar Consistency and Modality

In contrast to the other interfaces tested, Model C consisted of a tree of menus 3-5 levels deep. At any given time, the active grammar commands were based on the active node of the menu tree. In addition, the voice commands for the top two levels of the tree (including voice commands for navigation through the menu tree) were always active. Visually, up to three levels of the menu tree were displayed on a single screen.

The command language for Model C was inconsistent and hard to remember. The voice feedback for a command sometimes echoed the recognized command and sometimes told the user what action the system had taken. However, there was a visual display of the command recognized from the last voice utterance, which was useful. There was a good correspondence between the text of on-screen objects/icons and the command language used to select them, but there was no clear visual indication of which objects could be manipulated by voice.

The combination of menu complexity and grammar inconsistency made it difficult for the user to determine the active menu node and to infer what voice commands were available in a particular state. This system had a very high cost for recognition errors because a misrecognition would usually cause a transition to a state very far away in the menu tree, with no way to recover.

Summary of Evaluation of Current Interfaces

Model A was superior to the other interfaces that we tested. The consistency of the UI design and adherence to good design principles was the primary reason this was the best of the three tested systems, although the fact that this system had good ASR performance also contributed. In those areas of the system where the design had inconsistent grammars and/or visual cues, the resulting misrecognitions combined with the lack of rejection of out-of-grammar utterances to significantly degrade the usability. The 'back' functionality provided by this interface could have been a significant feature if it had worked consistently.

Model B was the worst of the interfaces tested. In spite of the poor ASR performance, the main weakness of this system was the implementation of the Push-and-Release with listening tone. This system provides an example of how an interface implementation problem can manifest as an ASR problem. Here, recognition performance was seriously degraded by the variable delay of the listening tone. In addition, the lack of rejection, and the relatively poor underlying ASR performance (especially in noise), all combined to make this interface unacceptably poor.

Model C had the most complicated modal structure. It also had system-paced turn-taking, which usually results in higher cognitive load on the user. Its speech recognition was reasonably good, with some notable exceptions.

In general, current state of the art systems show very distinct design philosophies. While we are in favor of innovative design, we hope that the design of future systems will more thoroughly consider the driver's cognitive load. It is interesting to note that the underlying speech recognition performance, while significant, was far from the most important factor in the usability of these interfaces—this shows the importance of good interface design in this application space.

A SPEECH-IN LIST-OUT APPROACH TO IN-CAR SPOKEN USER INTERFACES

Many in-car applications for which spoken UIs may be used deal with the selection of one of an enumerable set of possible responses, e.g., selecting one of a number of radio stations, retrieving a song from a music collection, selecting a point of interest (from the set of all points of interest), etc. The most common UI for these applications is through a hierarchy of menus. Even when the UI is speech-based, speech is used primarily as an input or output mechanism for the underlying menu-driven interface.

At Mitsubishi Electric Research Labs we have developed an alternative speech UI for selecting an element from a set, which we refer to as the "Speech-In List-Out" interface or SILO. In this section, we briefly describe the SILO interface. We also describe an experiment that indicates that the SILO interface can result in lower driving interference than the menu-driven interface.

Selection from a Set

The most common paradigm for retrieving a specific response from a large set is through menus; however, as the size of the selection set increases (e.g. for a UI to a digital music player with an ever increasing repertoire of songs), the tree of menus increases in depth and width, and can become problematic, particularly for users who are also simultaneously involved in other attention-critical tasks such as driving.

While the problem may be alleviated to some degree through voice output and spoken input (Leatherby & Pausch, 1992; Cohen et al., 2000), this by itself is not a solution. Spoken enumeration of menu choices cannot fully replace visual display; a deeply nested menu-tree, presented aurally, is very demanding in terms of cognitive load. Knowing that a quick glance at a screen can recover forgotten information relieves the user from having to keep close track of the system's state in their mind. Spoken-input-based interfaces must address the problem of misrecognition errors (e.g., in noisy environments) and, more importantly, the "what can I say" problem—users must be able to intuit what to say to the speech recognition system (which typically works from rigid grammars for such tasks). This latter issue can be particularly difficult when selecting from long lists.

The SILO interface (Divi et al., 2004) recasts the UI as a search problem. The set of all possible responses are viewed as documents in an index. The user prompts the system with a single spoken input, which is treated as a query into this index. The system returns a short *list* of possible matches to the query. The user must make the final selection from this list. While search-based speech UIs have previously been proposed (e.g., Cohen, 1991), SILO differs from them in that it places no restrictions on the user's language. The user is not required to learn a grammar of query terms. The simplicity of the resulting interactions between the user and the system is expected to result in a lower cognitive load on the user, an important consideration in the automotive environment.

The enabling technology for SILO is the SpokenQuery (SQ) speech-based search engine (Wolf & Raj, 2002). SQ is similar to text-based information retrieval engines except that users speak the query instead of typing it. Users may say whatever words they think best describe the desired items. There is no rigid grammar or vocabulary. The output is an ordered list of items judged to be pertinent to the query.

A major problem for speech UIs is misrecognition error, which can derail an interaction. The reason is that they attempt to convert the user's spoken input to an unambiguous text string, prior to processing. In contrast, SQ converts the spoken input to a set of words with associated probabilities, which is then used to retrieve documents from an index. The set of words and their associated probabilities are derived from the *recognition lattice* that represents candidate words considered by the recognizer. The lattice often includes the actual words spoken by the user even when they are not included in the disambiguated text output. As a result, SQ is able to perform well even in highly noisy conditions (such as automobiles) in which speech UIs that depend on accurate recognition fail. Consequently, SILO interfaces are able to perform robustly in noisy environments (Divi et al., 2004). Table 1 lists some example phrases and their (often poor) interpretation by the speech recognizer along with the performance of the SQ search.

Though the disambiguated phrase output by the speech recognition system is often wildly inaccurate, SQ manages to return the desired song near or at the top of the list. The right-most column shows the rank of the desired result in SQ's output.

The design of the SILO interface follows several of the design principles enumerated above. Specifically:

- **Interactions should be user paced:** The system always waits for the user to initiate the next interaction.
- **Appropriate use of speech:** Speech input is used only for choosing from very large sets where the use of buttons (for scrolling and selection) is inefficient or impossible. All choices from small sets are performed by direct manipulation.
- **Non-modal:** SILO is not modal, therefore the user does not need to remember the system state.

Experimental Evaluation of the SILO Interface

We designed an experiment to compare the SILO interface to a menu-based UI for in-car music selection. An effective in-car UI must not only allow users to find desired information quickly, but also affect their driving performance minimally. To evaluate both factors, we compared quantitative measurements of simulated steering and

Table 1. Example of SQ search to retrieve songs from a collection

User says…	**System hears…**	**SILO search result**
"Play Walking in my shoes by Depesh Mode"	layla [NOISE] issues [NOISE] [NOISE] load	1
"Depesh Mode, Walking in my shoes"	E [NOISE] looking [NOISE] night shoes	1
"Walking in my shoes"	law(2) pinion mae issues	1
"Walking in my shoes by Billy Joel" (partially incorrect information)	walking inn might shoes night billie joel	1
"um, uh, get me Credence Clearwater Revival… um… Who'll stop the Rain" (extra words)	fall(2) [UH] dead beat creedence clearwater revival [UM] long will stop it rains	1
"Credence Clearwater Revival, Who'll stop the Rain" (very noisy environment)	[NOISE] [COUGH] clearwater revival [COUGH] down [COUGH] [BREATH]	6

braking while searching for music with the two interfaces.

Our hypotheses were:

- **H1:** Subjects will more accurately track a moving target with a steering wheel while searching for songs using the Mediafinder SILO interface than while using the menu-driven interface.
- **H2:** Subjects will react faster to a braking signal while searching for songs using the SILO interface than while using the menu-driven interface.
- **H3:** Subjects will be able to find songs faster while using the SILO interface than while using the menu-driven interface while driving.

Experiments were conducted on a simple driving simulator, such as those in Beusmans et al. (1995) and Driving Simulators (2006), that mimicked two important facets of driving—steering and braking. The simulator had both a "windshield" and "in-dash" display. Subjects steered, braked, and controlled the interfaces with a steering wheel and gas and brake pedals. A microphone was placed on top of the main monitor. Steering was measured with a pursuit tracking task in which the subject used the wheel to closely frame a moving target (Strayer et al., 2001). The simulator recorded the distance in pixels between the moving target and the user-controlled frame 30 times a second. Braking was measured by recording subjects' reaction time to circles that appeared on screen at random intervals. Subjects were asked to only react to moving circles and to ignore stationary ones. Moving and stationary circles were equally probable.

We built two interfaces for this study. The first was a menu-based interface based on a sampling of currently available MP3 jukeboxes; the second was the SILO interface. Both interfaces ran on the same "in-dash" display and were controlled using buttons on the steering wheel. Both interfaces searched the same music database of 2124 songs by 118 artists, and both were displayed at the same resolution in the same position relative to the subject. Additionally, both interfaces displayed the same number of lines of text in identical fonts. Neither interface dealt with many of the controls needed for a fully functional in-car audio system, such as volume, power, and radio controls.

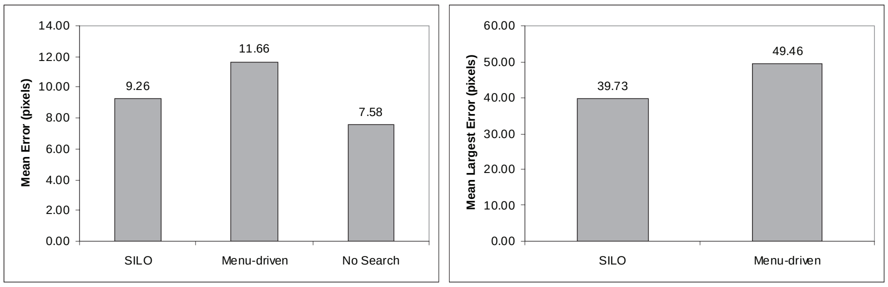
Fourteen subjects, eight male and six female, of ages ranging from 18 to 37, participated in this experiment. All but one were regular automobile drivers. Subjects were first instructed on how to correctly perform the steering and braking tasks and were given as much time as they wanted to practice "driving." Next, they were instructed to search for and playback specific songs while performing the driving task. Subjects completed 8 trials each with both the SILO and the menu-driven interfaces. Before each set of trials, subjects were instructed on how to use the current interface and allowed to practice searches while not driving. During each trial, the testing application displayed the steering and braking signals along with instructions asking the user to search for a specific song (e.g., "Please listen to the song Only the Good Die Young by Billy Joel from the album The Stranger"). Subjects were allowed to take a break between trials for as long as they wished. The order that the interfaces were used was balanced among participants, and the order of the requested songs was randomized. The application logged the distance between the moving target and the subject-controlled black box, as well as the reaction time to any brake stimulus presented during each trial. The task time was also logged, measured from the moment that the instructions appeared on the screen to the moment that the correct song started playing. To reduce learning effects, only the last 4 of each set of 8 trials contributed to the results.

Results

Our data supports hypotheses H1 and H3 and rejects H2.

- **H1:** Subjects were able to steer more accurately while searching for music using the SILO interface than with the menu-driven interface (on average, 9.2 vs. 11.6 pixels of

Figure 1. The SILO interface had both a significantly lower mean steering error (below left) and a significantly lower mean largest steering error (below right) than the menu-driven interface (Source: Forlines et al., 2005)



error respectively, t(13) = 3.15, p=0.003). However subjects steered most accurately while driving without searching (on average, 7.4 vs. 9.2 pixels for SILO, t(13)=2.5, p=0.013). The average error for each condition is shown in Figure 1 (*below left*). The SILO interface had a significantly lower *maximum* steering error as well (39.7 pixels vs. 49.4 pixels, t(13)=2.27, p=0.02). This measurement of error roughly corresponds to the point when the subject was most distracted from the steering task. If actually driving, this point would be the point of greatest lane exceedence. The average maximum error for the two interfaces is shown in Figure 1 (*below right*).

- **H2:** The mean breaking reaction times were indistinguishable between the SILO and menu-driven conditions (on average, 1196 ms vs. 1057 ms, t(13)=1.66, p=0.12); however, subjects were significantly faster at braking while not searching for music than while searching using the SILO (p=0.008) or the menu-driven (p=0.03) interface. The mean reaction time to the brake stimulus for each condition is shown in Figure 2.
- **H3:** Subjects were significantly faster at finding and playing a specific song while using the SILO interface than while using the menu-driven interface (on average, 18.0 vs. 25.2 sec., t(13)=2.69, p=0.009). The mean search time for each interface is shown in Figure 3. It is important to note that it was not unusual for the SILO interface to have a computational interruption of 3-6 seconds, which was included in the SILO search time. A faster CPU or better microphone could decrease this time.

Figure 2. There was no significant difference in mean break reaction times between the search conditions (Source: Forlines et al., 2005)

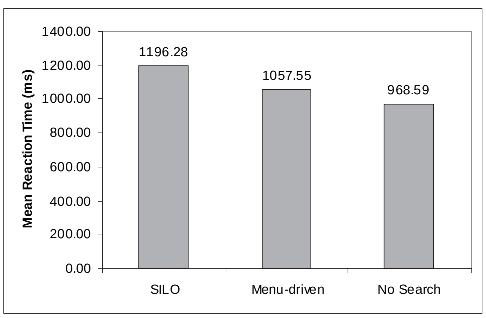
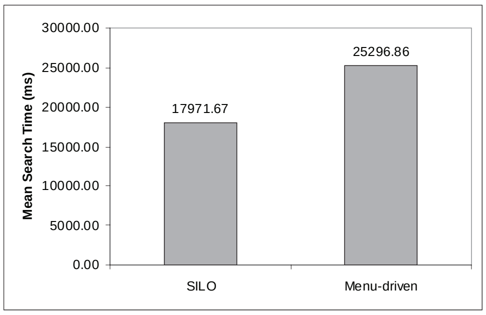


Figure 3. Subjects were significantly faster at finding songs with the SILO interface (Source: Forlines et al., 2005)



Since SILO returns a list of songs from which the user must select, an important factor is the position of the correct song in the returned list. In 35 out of the 56 trials SILO returned the correct song at the top of the list on the first try. The average position for the correct song for all SILO trials was 5.1.

Experimental Discussion

The evidence indicates that our SILO interface for music finding has measurable advantages over the standard menu-based approach for users operating a simulated automobile. The SILO interface scores better than the menu-based system on several of the design goals mentioned outlined above. It poses lower cognitive load, as evidenced by the improved driving accuracy, has lower interaction time, and is more effective for the experienced user. Although we observed no statistical difference in mean break reaction times between the SILO and menu-driven interfaces, closer inspection revealed that subjects were less likely to encounter a brake stimulus while using SILO due to the faster task completion. SILO effectively results in fewer opportunities for braking error.

The SILO interface has several additional advantages that were not explicitly evaluated in the experiment. We expect that these will be subjects of future study:

- **Flexibility:** The SILO interface was able to retrieve songs from only partial information, such as the song title. On the other hand, for the menu-based interface, it would not have been sufficient to ask subjects to find the song “Never Die” without telling them it is by the artist “Creed” on the album “Human Clay.”
- **Scalability:** The song library used for the study contained only 2124 songs. The latest handheld music players can hold over 10,000 songs. As the number of available artists, albums, and songs grows, we expect the time needed to search through a menu-driven interface to grow as well. An informal evaluation of the SILO interface searching a database of 250,000 songs shows no noticeable differences in search time.
- **Robustness:** The metadata in the music files in our library was not always consistent. For example, music by the group “The B-52s” was erroneously split into many artists: “The B-52s,” “B-52s,” “The B52s,” etc. While these inconsistencies were problematic for the menu-driven interface, they do not affect the SILO interface.

The experiments reported here only evaluated SILO in a limited music selection task. Many music jukeboxes can present their content in alternative fashions such as user defined playlists, favorites, etc. Mediafinder is easily modifiable to handle playlists and personalization; however a rigorous evaluation of its capabilities in this direction remains to be performed.

Other limitations of this study include the fact that an actual in-car environment that included environmental noise was not used. Other tests using automotive speech data have shown that the SpokenQuery information retrieval engine is very robust to high levels of environmental noise (e.g., Divi et al., 2004). We are therefore optimistic about the performance of the SILO interface in real in-car environments, but must confirm this expectation with future experiments. Finally, we look forward to a comparison between SILO and other speech based music selection systems.

FUTURE TRENDS

The current trend toward increasing the number and complexity of secondary tasks for automobile drivers is both worrisome and accelerating. The risks associated with these complex tasks raise two opportunities. First, standards bodies, government agencies, and the public must demand less distracting, safer interfaces for drivers. Second, manufacturers must provide these interfaces, and perhaps market not only their additional functionality, but also their safety advantages. Researchers and manufacturers must conduct studies measuring the degree of interference between the driving task and existing and future automotive user interfaces. Safety and liability dictate that these experiments be conducted in a driving simulator environment.

CONCLUSION

The advent of complex user interfaces in automobiles raises many issues relating to safety and usability, some of which can be mitigated by the appropriate use of speech in the UI. We have presented a set of design principles that can help mitigate some of the problems cited, and have applied these principles in a review of several existing automotive speech interfaces. Finally, we presented an in-car interface for the selection of items from a large collection and have shown that this method interferes less with driving than the current status quo. Much work remains to maximize the safety and usability of complex devices in automobiles, and we hope that this writing will aid the UI designer in this endeavor.

REFERENCES

Beusmans, J., & Rensink, R. (Eds.) (1995). *Cambridge Basic Research 1995 Annual Report* (Tech. Rep. No. CBR-TR-95-7). Cambridge, MA: Nissan Cambridge Basic Research.

Cohen, P. (1991). Integrated interfaces for decision-support with simulation. In *Proceedings of the 23rd Winter Conference on Simulation* (pp. 1066-1072).

Cohen, P., McGee, D., & Clow, J. (2000). The efficiency of multimodal interaction for a map-based task. In *Proceedings of the Applied Natural Language Processing Conference (ANLP'00)* (pp. 331-338). Seattle, WA: Morgan Kaufmann.

Divi, V., Forlines, C., Van Gemert, J.V., Raj, B., Schmidt-Nielsen, B., Wittenburg, K., Woelfel, J., Wolf, P., & Zhang, F. (2004, May 2-7). A speech-in-list-out approach to spoken user interfaces. In *Proceedings of the Human Language Technology Conference (HLT 2004)* (pp. 112-116). Boston, MA.

Driving Simulators. (2006). Retrieved October 24, 2006 from http://www.inrets.fr/ur/sara/Pg_simus_e.html

Forlines, C., Schimdt-Nielsen, B., Raj, B., Wittenburg, K., & Wolf, P. (2005). A comparison between spoken queries and menu-based interfaces for in-car digital music selection. *IFIP TC13 International Conference on Human-Computer Interaction (INTERACT 2005)*

Klauer, S.G., Dingus, T. A., Neale, V. L., Sudweeks, J.D., & Ramsey, D.J. (2006). *The Impact of Driver Inattention on Near-Crash/Crash Risk: An Analysis Using the 100-Car Naturalistic Driving Study Data*. (DOT HS 810 594) Washington, DC: United States Government National Highway Traffic Safety Administration, Office of Human-Vehicle Performance Research.

Leatherby, J.H., & Pausch, R. (1992) Voice input as a replacement for keyboard accelerators in a mousebased graphical editor: An empirical study. *Journal of the American Voice Input/Output Society, 11*, 2.

Ranney T., Watson, G. S., Mazzae, E. N., Papelis, Y. E., Ahmad O., & Wightman, J. R. (2004). *Examination of the distraction effects of wireless phone interfaces using the national advanced driving simulator—Preliminary report on freeway pilot study* (DOT 809 737). East Liberty, OH: United States Government National Highway Traffic Safety Administration, Vehicle Research and Test Center

Society of Automotive Engineers. (2004). *SAE recommended practice navigation and route guidance function accessibility while driving* (SAE 2364). Warrendale, PA: Society of Automotive Engineers.

Strayer, D.L., Drews, F. A., Albert, R. W., & Johnston, W. A. (2001). Cell phone induced perceptual impairments during simulated driving. In D. V. McGehee, J. D. Lee, & M. Rizzo (Eds.), *Driving assessment 2001: International symposium on human factors in driver assessment, training, and vehicle design*.

Stutts, J.C., Reinfurt, D.W., Staplin, L.W., & Rodgman, E.A. (2001). *The role of driver distraction in traffic crashes*. Washington, DC: AAA Foundation for Traffic Safety. Retrieved October 24, 2006, from http://www.aaafts.org/pdf/distraction.pdf.

US Census Bureau News. (2005). Americans spend more than 100 hours commuting to work each year. *Census Bureau Reports*, Retrieved October 24, 2006, from http://www.census.gov/Press-Release/www/releases/archives/ american_community_survey_acs/004489.html.

UMTRI Driver Interface Group. (updated 2006). *All the major telematics guidelines we've seen*. University of Michigan Transportation Research Institute, Driver Interface Group. Retrieved October 26, 2006, from http://www.umich.edu/~driving/guidelines/guidelines.html

Wang, J., Knipling, R. R., & Goodman, M. J. (1996). The role of driver inattention in crashes; new statistics from the 1995 crashworthiness data system (CDS). In *proceedings of the 40th Annual Proceedings: Association for the Advancement of Automotive Medicine* (pp. 377-392).

Wolf, P., & Raj, B. (2002). The MERL Spoken-Query Information Retrieval System: A System for Retrieving Pertinent Documents from a Spoken Query. In *Proceedings of ICME, Vol. 2* (pp. 317-320).

KEY TERMS

Cognitive Load: A measure of the mental effort required to carry out a given task.

Driver Distraction: A measure of the degree to which attention is taken away from the driving task.

Listening Tone: A sound generated by a speech-based user interface when it is ready to accept spoken input

Lombard Effect: The specific changes in style of speech caused by the presence of noise. In particular the speech gets louder and higher frequencies are emphasized

Misrecognition: A speech recognition result which does not accurately represent what was spoken by the user. In spoken command recognition, recognizing the exact words spoken is not necessary to avoid a misrecognition as long as the correct command is recognized.

Push and Hold: A type of speech interaction where the user must hold down a button while speaking to the system. This kind of system is familiar to most users as it is reminiscent of a walkie-talkie.

Push and Release: A type of speech interaction where the user must depress a button prior to the start of speech. This type of interaction is unfamiliar to some users, but provides an easy learning curve with the proper affordances.

Recognition lattice: A directed graph of candidate words considered by a speech recognizer. This graph will often contain alternate words with similar phonetics. It will also contain confidence weights.

SILO: A speech-based user interface which returns a shortlist of possible responses, from which the user must make a final selection. We refer to such interfaces as Speech-In List-Out, or SILO.

Speech-Based [User] Interface (SUI): A user interface which uses utterances spoken by the user as a primary input mode. A speech based interface may also have other input modes, such as dedicated or softkey input, and may also have voice feedback and/or visual feedback.

Telematics: Broadly, telematics refers to the combination of telecommunication and computation. More specifically telematics has come to refer to mobile systems which combine wireless data communications with local computation resources. Voice communication and/or location information provided by GPS are often assumed.

Chapter XVI
Design for Mobile Learning in Museums

Nikolaos Tselios
University of Patras, Greece

Ioanna Papadimitriou
University of Patras, Greece

Dimitrios Raptis
University of Patras, Greece

Nikoletta Yiannoutsou
University of Patras, Greece

Vassilis Komis
University of Patras, Greece

Nikolaos Avouris
University of Patras, Greece

ABSTRACT

This chapter discusses the design challenges of mobile museum learning applications. Museums are undoubtedly rich in learning opportunities to be further enhanced with effective use of mobile technology. A visit supported and mediated by mobile devices can trigger the visitors' motivation by stimulating their imagination and engagement, giving opportunities to reorganize and conceptualise historical, cultural and technological facts in a constructive and meaningful way. In particular, context of use, social and constructivist aspects of learning and novel pedagogical approaches are important factors to be taken in consideration during the design process. A thorough study of existing systems is presented in the chapter in order to offer a background for extracting useful design approaches and guidelines. The chapter closes with a discussion on our experience in designing a collaborative learning activity for a cultural history museum.

INTRODUCTION

Use of mobile devices spreads in everyday human activities. These devices offer portability, wireless communication and connectivity to information resources and are primarily used as mobile digital assistants and communication mediators. Thus, it is no surprise that various attempts to use mobile appliances for learning purposes have been reported either inside or outside school (Roschelle, 2003). The term *mobile learning* or *m-learning* has been coined and concerns the use of wireless technologies, portable appliances and applications in the learning process without location or time restrictions. Practitioners' reports (Perry, 2003; Vahey & Crawford, 2002) and scientific findings (Norris & Soloway, 2004; Roschelle, 2003; Zurita & Nussbaum, 2004) communicate promising results in using these applications in various educational activities. The related bibliography proposes various uses of mobile appliances for learning. These Activities might concern access and management of information and communication and collaboration between users, under the frame of various learning situations.

A particular domain related to collaborative learning is defined as the support provided towards the educational goals through a coordinated and shared activity (Dillenbourg, 1999). In such cases, peer interactions involved as a result of the effort to build and support collaborative problem solving, are thought to be conducive to learning. On the other hand, traditional groupware environments are known to have various technological constraints which inflict on the learning process (Myers et al., 1998). Therefore, mobile collaborative learning systems (mCSCL) are recognized as a potential solution, as they support a more natural cooperative environment due to their wireless connectivity and portability (Danesh, Inkpen, Lau et al., 2001). While the mobility in physical space is of primary importance for establishing social interaction, this ability is reduced when interacting through a desktop system. It is evident that, by retaining the ability to move around it is easier to establish a social dialogue and two discrete communication channels may be simultaneously established through devices: one physical and one digital. Additionally, a mobile device can be treated as an information collector in a lab or in an information rich space (Rieger & Gay, 1997), as a book, as an organizing medium during transportation or even as a mediation of rich and stimulating interaction with the environment (i.e., in a museum). Effective usage of mobile appliances has been reported in language learning, mathematics, natural and social sciences (Luchini et al., 2002).

Furthermore, various technological constraints need to be taken in consideration during the design of activities which involve mobile devices. Such an example is the small screen, which cannot present all the information of interest while the lack of a full keyboard creates constraints in relation to data entry (Hayhoe, 2001). There is a need to provide the user with the possibility to 'go large' by getting information from both the virtual and physical world, while simultaneously 'going small,' by retrieving the useful and complementary information and getting involved into meaningful and easy to accomplish tasks (Luchini et al., 2002). In addition, despite the fact that technological solutions are proliferating and maturing, we still have a partial understanding of how users take effectively advantage of mobile devices. Specifically, in relation to communication and interaction, we need to investigate how mobile technology can be used for development of social networks and how it can provide richer ways for people to communicate and engage with others. In public spaces, like museums, a crucial question is if the serendipitous exchanges and interactions that often occur should be supported through mobile technology, how and where the interaction between people takes place and how is affected by this novel technology. Clearly, a better understanding of social activities and social interactions in public spaces should emerge to answer these questions.

A number of the aforementioned issues are discussed next in the context of a museum visit. First, we analyse how the context can affect any activity and application design. Then, we outline the most promising mobile learning applications and finally, we present our experiences of introducing collaborative learning activities using a novel approach based on the best practices surveyed previously, in a large scale project for a cultural history museum.

INTERACTION DESIGN FOR MOBILE APPLICATIONS

Interaction design is one of the main challenges of mobile applications design. Direct transfer of knowledge and practices from the user-desktop interaction metaphor, without taking in consideration the challenges of the new interaction paradigm is not effective. A new conceptualization of interaction is needed for ubiquitous computing. The traditional definition (Norman, 1986) of the user interface as a "means by which people and computers communicate with each other," becomes in ubiquitous computing, the means by which the people and the environment communicate with each other *facilitated* by mobile devices. As a result, interaction design is fundamentally different. In the traditional case, the user interacts with the computer with the intention to carry out a task. The reaction of the computer to user actions modifies its state and results in a dialogue between the human and the machine.

On the other hand, the user interaction with mobile devices is triadic, as the interaction is equally affected not only by the action of the user and the system's response, but by the context of use itself. The level of transparency of the environment, taking into account the presence of the mobile device and the degree of support to 'environmental' tasks meaningful to the user, are new issues to be considered. Consequently, new interaction design and evaluation criteria are required, since the design should not only focus on the user experience but pay also attention to the presence of other devices or objects of interest, including the level of awareness of the environment. By building the virtual information space into the real, the real is enhanced, but conversely, by drawing upon the physical, there is the opportunity to make the virtual space more tangible and intuitive and lower the overall cognitive load associated with each task.

To summarize, a number of design principles are proposed for mobile applications design:

a. Effective and efficient *context awareness* methods and models, with respect to the concept of context as defined by Dey (2001): 'Context is any information that can be used to characterize the situation of an entity. An entity is a person, place, or object that is considered relevant to the interaction between a user and an application, including the user and applications themselves.'
b. Presentation of useful information to the user *complementary* to the information communicated by the environment.
c. A*ccurate and timely update* of environmental data that affect the quality of interaction.
d. Contextualized and personalized information according to *personal needs.*
e. Information should be *presented to the user* rather than having the user searching for it.

Failure to look into these design issues can lead to erroneous interaction. For example, delays of the network, lack of synchronization between two artifacts of the environment or slight repositioning of the device can lead to misconceptions and illegal interaction states. In addition, information flow models should be aligned according to the information push requirement and relevant user modeling and adaptation techniques to support this flow of information should be defined. Finally, new usability evaluation techniques, concerning mobile applications should emerge to shape a novel interaction paradigm.

Dix et al. (2000) present a framework to systematically address the discussed design issues and successive context awareness elements are inserted in the design process: (a) the *infrastructure* level (i.e., available network bandwidth, displays' resolution), (b) the *system* level (type and pace of feedback and feed through), (c) the *domain* level (the degree of adaptability that a system must provide to different users) and (d) the *physical* level (physical attributes of the device, location method and the environment). All these elements should be tackled independently and as a whole in order to study the effect of every design decision to each other.

We formulate the interaction design aspects discussed through the problem of designing a mobile learning application for a museum. During the visit a user has only a partial understanding of the available exhibits. This situation can be supported by complementary information included in the physical environment, for example alternative representations, concerning the historical role of the people or the artifacts presented the artistic value of a painting (Evans & Sterry, 1999), and so forth. This cognitive process of immersion into the cultural context, represented by the museum exhibits, could be supported by drawing upon the stimuli produced during the visit using context aware mobile devices. Therefore, these devices should be viewed as tools to enhance the involvement of a user in the cultural discovery process, tools that challenge the user to imagine the social, historical and cultural context, aligning her to a meaningful and worthy experience.

It is not argued here that the infusion of mobile technologies in museums will necessarily result to meaningful learning processes. Our analysis involves the potential use of the technology when integrated in educational activities (Hall & Bannon, 2006), which will offer a structured learning activity according to the characteristics of museums' content and the functionalities of the technologies used. To better illustrate this point (a) we briefly present a set of selected exemplary cases which demonstrate different ways of integrating mobile educational applications in museums and (b) we provide a more detailed account of such an application that we designed for a museum in Greece.

In the next section a number of approaches supporting such a visit are reviewed and examined using the design aspects as guiding paradigm and point of reference. Since the goal of the visitor is to see and learn more and not to explicitly use technology, a deep understanding of visitors' needs is important during the design phase, to avoid disturbances that can destruct her from her objective. Therefore, decisions made for the technology used and the styles of interaction, with the involved devices, have to deal with user's patterns of visit. Having the above requirements in hand, we use the framework proposed by Dix et al. (2000) to organize a coherent characteristics inspection of some representative examples of mobile museum guides.

MOBILE DEVICES AS MUSEUM GUIDES

In this section, some representative design approaches for mobile museum applications are discussed. An extended survey is included in Raptis, Tselios and Avouris (2005). The first system named "Electronic Guidebook," deployed in the Exploratorium science museum (Fleck et al., 2002), tries to involve the visitors to directly manipulate the exhibits and provides instructions as well as additional science explanations about the natural phenomena people are watching. The system of the Marble Museum of Carrara (Ciavarella & Paterno, 2004) stores the information locally in the PDA's memory, uses a map to guide the visitors around the museum and presents content of different abstraction levels (i.e., room, section and exhibit). The "ImogI" system uses Bluetooth to establish communication between the PDAs and exhibits and presents the closest exhibits to the user, (Luyten & Coninx, 2004). The "Sotto Voce" system gives details about everyday things located in an old house (Grinter et al., 2002) by having pictures of the walls on the PDA's screen and asking from the user to select the exhibit she is interested in, by pressing it. The "Points of Departure" system (www.sfmoma.org) gives details in video and audio form by having 'thumbnails' of exhibits on the PDA's screen. It also uses 'Smart Tables' in order to enrich the interaction. A system, in the Lasar Segall Museum, Sao Paolo, Brazil (Dyan, 2004), automatically delivers information to the PDA, about more than 3,000 paintings. In the Tokyo University Digital Museum a system uses three different approaches to deliver content. The

PDMA, in which the user holds the device above the exhibit she is interested in, the Point-it, in which the visitor uses laser-pointer to select specific exhibits and finally the Museum AR in which visitors wear glasses in order to get details about the exhibits (Koshizuka & Sakamura, 2000).

The system developed in the C-Map project, (Mase, Sumi, & Kadobayashi, 2000), uses active badges to simulate the location of the visitor, allowing tour planning and a VR system, controlled by the gestures of the visitor. In a Tour guide (Chou, Lee, Lee et al., 2004), the information about the exhibits is automatically presented and there is no variation in the form of the visit, but subjective tour guides are used. A different approach is the one adopted in the Museum of Fine Arts in Antwerp (Van Gool, Tuytelaars, & Pollefeys, 1999), in which the user is equipped with a camera and selects exhibits, or details of an exhibit by taking pictures. A tour guide in the PEACH project, (Rocchi, Stock, Zancanaro et al., 2004), which migrates the interaction from the PDA to screens and uses a TV-like metaphor, using 'newscasters' to deliver content. Finally, a nomadic information system, the Hippie, developed in the framework of the HIPS project, (Oppermann & Specht, 1999), allows the user to access a personal virtual space during or after the visit. In the latter system, an electronic compass is used to identify the direction of a visitor.

The *infrastructure* context concerns the connections between the devices that comprise the system and influence the validity of the information that is provided through them to the users and needs not only to be addressed in problematic situations. It is also related with the validity and timely updates of available information. This can be clearly seen in collaboration activities where

Table 1. Design decisions affecting system context

	Location technology	**Storage of information**	**Flow of information**	**Additional functions**
"Antwerp project"	IrDA	In Server	Active	Cameras
C-Map	IrDA	In Server	Active, exhibit recommendations	Active Badges, Screens
Hippie	IrDA	In Server	Active, info presented based on the history of visit	
ImogI	Bluetooth	Info stored in Bluetooth transmitters	Active, proximity manager	
Lasar Segal Museum	IrDA	In Server	Passive	
Marble Museum	IrDA	Locally stored info, abstraction levels	Active, history of the visit	
PDMA, Point it, Museum AR	IrDA	In Server	Active	laser pointer, glasses
PEACH project	IrDA	In Server	Passive, task migration	Screens
Points of departure		Locally stored info	Active	Screens
Rememberer	RFID	In Server	Passive	Cameras
Sotto Voce		Locally stored info	Active	
Tour Guide System (Taiwan)	IrDA	In Server	Passive, subjective tour guides	

the user constantly needs to know the location of other users, the virtual space, the shared objects, and so forth. In the specific museum domain the results may not be so critical but can lead the user to various misunderstandings.

The mentioned systems use an indirect way of informing the user that her requests have been carried out: the user sees and hears the reflection of her requests on the PDA. There is no clear notification that the user's demands are executed successfully or not. Some of the systems use external factors, as signs of success, such as a led light ("Rememberer") and audio signals ("Marble Museum"). But in general terms, the user is on his or her own when problems occur and the systems leave it up to her to find it out, by observing that, there is no progress. We have to point that it could be very distracting and even annoying to have feedback messages in every state of interaction, but it is important for designers, to include a non-intrusive approach to inform that there is a problem and provide constructive feedback to overcome it.

Regarding the *system* context we can distinguish four different approaches as a means of awareness technology. In the first approach (Table 1), the PDA is the whole system. There are no other devices or awareness mechanisms involved and the information presented to the user is stored locally in the PDA. The second approach uses RFID tags to establish communication between the PDAs and the exhibits and the third which uses Bluetooth to establish communication with the exhibits and deliver content. The forth and most common approach uses IrDA technology to estimate the position of the visitor in space. Usually, IrDA tags are placed near every exhibit or in the entrance of each exhibition room and Wi-Fi derives the information to the PDA from a server Also, many different additional devices are built and integrated into these systems like screens (as a standalone devices or as interacting devices with the PDAs, where the user has the opportunity to transmit sequentially her interaction with the system from the PDA to a Screen). Regarding the *location*, all the studied systems use a topological approach to identify the position of a PDA, which informs approximately the system about the user's location. However, in the case of a museum with densely place exhibits, a more precise Cartesian approach can yield accurate user localization.

Domain context concerns aspects related to the situated interaction that takes place in the specific domain. Often in museum applications there is a lack of information about user profiles and characteristics. It is however important to consider that each visitor in a museum has different expectations, and is interested in different aspects regarding the exhibits. In the studied systems only in those that allow interaction of the users with servers there is a possibility for personalized interaction. Most of the systems require from the user to login, answer some specific questions, in order to build a model of the user and present the information in her PDA according to her language, her expertise level and her physical needs (i.e., bigger fonts for those with sight impediments). When domain context is absent from the design process the system operates as a tool suited for the needs of a single hypothetical 'ideal' user. In such an environment this 'ideal' user will likely represent the needs and expectations of a small fraction of real visitors.

The system may push information to the user or it may wait until the user decides to pull it from the system. In the first case, special consideration should be taken to the user's specific activity and objective. Questions related to situated domain context are the following: Does the system propose any relevant information based on the history of users interaction? Does it adapt to actions repeatedly made by the user? Does it present content in different ways? For example, the "ImogI" system rearranges the order of the icons putting in front the mostly used ones. Also, in PEACH and in 'Points of Departure' the user can change the interaction medium from PDAs to Screens, in order to see more detailed information.

The *physical* context lays in the relation of the system with the physical environment and in problems concerning the physical nature of the devices. However, in the studied systems there is not a single mechanism of identifying the physical conditions. For example, in a room full with people, where a lot of noise exists, it would be appropriate if the system could automatically switch from an audio to a text presentation.

From the survey of the mobile guides applications presented here it seems that efficient design approaches could be achieved by augmenting physical space with information exchanges, by allowing collaboration and communication, by enhancing interactivity with the museum exhibits and by seamlessly integrating instantly available information delivered in various forms. However, the synergy between technology and pedagogy is not straightforward especially if we take into account the need to tackle issues such as efficient context integration, transparent usage of the PDA, and novel pedagogical approaches to exploit the capabilities of mobile devices. As a result, after discussing in detail usages of a mobile device as a mean of museum guidance, in the following, we attempt to discuss explicit educational activities mediated by mobile devices and a specific example of a new Mobile Learning environment.

DESIGNING MUSEUM MOBILE EDUCATIONAL ACTIVITIES

The level of exploitation of mobile devices in a museum setting is increasing and part of this use may have educational value. In this section we will focus on the added value of integrating educational mobile applications in museums. We will start our analysis by posing two questions that we consider central to this issue: (a) what is changing in the learning process taking place in a museum when mediated by mobile technology and (b) why these changes might be of educational or pedagogical interest? We will attempt to address these questions by focusing on three aspects related not only to the characteristics of mobile technology but also to the results of its integration in a museum. Specifically we will discuss: (a) the types of interaction between the visitor and the learning environment (e.g., the museum), (b) the learning activities that these interactions can support and (c) the role of context and motion in learning.

One facet of the learning process when mediated by mobile technology in museum visits involves the tangibility of museum artifacts: distant museum exhibits that were out there for the visitors allowing them just to observe now can be virtually touched, opened, turned and decomposed. In this case, technology provides to the user the key to open up the exhibit, explore it and construct an experience out of it. The traditional reading of information and observation of the exhibit is considered as one-dimensional "information flow" from the exhibit to the user. Mobile technology facilitates the transformation of the one dimensional relationship to a dialectic relationship between the user and the exhibit. Furthermore, this relationship can now include another important component (apart of the exhibits) of the museum environment: the other visitors. By providing a record of user–exhibit interaction for other visitors to see, reflect upon and transform technology can support social activities of communication, co-construction, and so forth between the visitors. To sum up, mobile technology mediates three types of interaction between the learner and the learning environment of a museum: (a) "exhibit–user" interaction (b) "user –exhibit" interaction and (c) "between the users" interaction about "a" and "b."

The enrichment of interaction between the learner and the museum might result in more or different learning opportunities (Cobb, 2002) the characteristics of which are outlined here. Specifically, the dialectic relationship between the user and the museum artifacts, mediated by mobile devices, might offer chances for analysis of the exhibit, experimentation with it, hypothesis formulation and testing, construction of interpretation, information processing and organization, reflection and many more, according to the educational activity designed. Collaboration and communication about the exhibits and information processing about them makes possible socio-constructive learning activities. By comparing these elements of the learning process to the reading or hearing of information about the exhibits (which is a the starting point for a non technology mediated museum visit) we realize that mobile technology has the potential to offer an active role to the learner: she can choose the information she wants to see, open up and de-construct an exhibit if she is interested in it, see how other visitors have interacted with a certain exhibit, discuss about it with them, exchange information, store information for further processing and use and so on.

Up until now, we described the role of mobile technology in learning with respect to two characteristics of the museum as learning environment: the exhibits and the other visitors. Another characteristic of the museum, which differentiates it from other learning environments (e.g., classroom) is that learning in a museum takes place while the learner moves. Learning while moving, quite often takes place very effectively without the support of technology. However there are cases that further processing with appropriate equipment is needed or some structuring of this "mobile learning experience" is proved to be useful. Mobile technologies can find in museums an important area of implementation not only because museum visits are structured around motion but because we have to support visitors *during* and not just after or before the visit (Patten, Arnedillo Sanchez et al., 2006). But why is it important to support learning during the visit? The answer here comes from the theory of situated learning (Lave & Wenger, 1991) which underlines the role of context in learning. Specifically, context facilitates knowledge construction by offering the practices, the tools, and the relevant background along with the objectives towards which learning is directed and has a specific meaning or a special function (knowledge is used for something). Finally, the use of mobile devices provides a new and very attractive way of interacting with the museum content especially for young children (Hall & Bannon, 2006).

As mentioned previously a large number of mobile applications have been developed during previous years for use in the museums (Raptis et al., 2005). All these mobile applications can add educational value to a museum visit in various ways. A survey of mobile educational applications for use inside the museum, led us to a categorization according to the educational approach followed in every occasion. The first category includes applications that mainly deliver information to the visitor and concerns the vast majority of applications created for museums. Mobile devices take the place of the museums' docents and offer predetermined guided tours based upon certain thematic criteria. The aforementioned applications offer the museum visitor an enhanced experience which can support the learning process through a behaviorist approach. Enhancement is succeeded by supplying multimedia and context-related content.

The second category of applications, suitable for educational use in museums, consists of applications which provide tools that can support the learning process in a more profound way. Compared to the first category, they provide information about the exhibits of a museum but furthermore they include a series of functions that increase the interactivity with the user. Such an example is the Sotto Voce System (Grinter et al., 2002), which includes an electronic guide with audio content and the ability of synchronized sharing of this content between visitors. Thus, the users can either use individually the guide or "eavesdrop" to the information that another visitor listens.

Another example is the applications developed for the Exploratorium, a science museum in San Francisco (Fleck et al., 2002). In this museum, the visitor has the possibility to manipulate and experiment with the exhibits. Also, an electronic guidance was designed to provide information about the exhibits and the phenomena related with them, posing relative questions to provide deeper visitors' engagement. These applications are closer to social-cultural learning theories as they provide the user with tools to organize and control the provided information.

The third category of educational applications presents a specific educational scenario. Usually, game-based activities where the users, mostly children aged 5-15, are challenged to act a role and complete carefully designed pedagogical tasks. Such an example is the MUSEX application (Yatani, Sugimot & Kusunoki, 2004), deployed in the National Museum of Emerging Science and

Innovation in Japan. MUSEX is a typical drill and practice educational system in which children work in pairs and are challenged to answer a number of questions. Children select an exhibit with their RFID reader equipped PDA and a question is presented in the screen with four possible answers. The activity is completed when each pair collects twelve correct answers. Children may collaborate and communicate either physically or via transceivers and monitor each group progress through a shared screen. After the completion of the activity the participators have the possibility to visit a Website and track their path inside the museum. The users can deeply interact with the exhibits, review the progress of her partner or ask for help (Yatani et al., 2004).

DinoHunter project includes several applications for the transmission of knowledge through game-based and mixed reality activities in the Senckenberg museum, Frankfurt, Germany. Three of these applications, namely DinoExplorer, DinoPick and DinoQuiz, are being supported by mobile technologies (Feix, Gobel, & Zumack, 2004). DinoExplorer delivers information to the users as an electronic guide, DinoPick allows the users to pick one part of the body of a dinosaur and get more multimedia information about this specific part and DinoQuiz provides a set of questions for further exploration of the exhibits of the museum.

Mystery at the Museum is another mobile, game-based, educational activity created for the Boston Museum of Science. It engages visitors in exploring and thinking in depth about the exhibits, thus making connections across them and encourages collaboration (Klopfer, Perry, Squire et al., 2005). High School students and their parents are called to solve a crime mystery where a band of thieves has stolen one of the exhibits. The users try to locate the criminals by using a PDA and a walkie-talkie. The participants must select upon the role of a technologist, a biologist or a detective. Depending on the chosen role they can interview virtual characters, pick up and examine virtual objects by using virtual equipment (e.g., microscope), collect virtual samples via infrared tags or exchange objects and interviews through the walkie-talkies. A study confirmed deep engagement of the participants and extensive collaboration due to the roles set.

Another similar approach is presented through the Scavenger Hunt Game activity used in the Chicago Historical Society Museum (Kwak, 2004). In this case, the children are challenged to answer a series of questions related to the exhibits and the local history. They undertake the role of a historical researcher and they are called to answer 10 multiple choice questions while examining the exhibits. Each user is individually engaged into the activity and her progress is evaluated in a way similar to electronic games. The Cicero Project implemented in the Marble Museum of Carrara introduces a variety of games to the visitors (Laurillau & Paternò, 2004). The games vary from finding the missing parts of a puzzle to answering questions about the exhibits. Its main characteristic is the support it provides to the visitors to socially interact and collaboratively participate in activities concerning the exhibits of the museum, through peer-awareness mechanisms.

A series of mobile educational activities was also carried out in the frame of the Handscape Project in the Johnson Museum (Thom-Santelli, Toma, Boehner et al., 2005). The “Museum Detective” engage students in role-playing activities. Children working in pairs are called to locate an object described by one clue and learn as much as possible for it. A series of multiple-choice questions is presented for further exploration of the exhibit. Four types of interactive element are also provided for the exhibits: a painting, a drawing activity and a building activity and a multimedia narrative. The multiple-choice questions and the building activity were drill and practice activities and the rest were activities allowing children to make their own creations.

The systems of the latter category present coherent learning experiences comprised of planned and organized pedagogic activities, where an intervention has been purposefully designed to result a positive impact on children’s cognitive and affective development. With respect to the contextual and interaction issues presented in the previous sections, we attempt to present in the next section an integrated application that involves children as role-playing characters by exploring the museum using a PDA.

AN EXAMPLE OF MOBILE ACTIVITY DESIGN FOR INDOOR MUSEUM VISIT

The "Inheritance" activity discussed here, is designed to support learning in the context of a cultural/historical museum visit. The application involves role-playing, information retrieval, data collecting and collaboration educational activities, suitable for children aged 10 or above working in teams of two or three members each. The activity scenario describes an imaginary story where the students are asked to help the Museum in finding the will of a deceased historian, worked for years in it. This will is hidden behind the historians' favorite exhibit. Clues to locate the document are scattered among the descriptions of some exhibits. If the children manage to find the will, all of the property of the historian will be inherited by the museum and not by his "greedy relatives." The scenario urges the students to read the description of the exhibits, find the clues and collaboratively locate the specific one.

During the design process of the activity we had to study the *museum context*, the *mobile technology* used and the *learning approach* to be followed in order to achieve the desired pedagogical outcome. The survey discussed in the previous section led us to adopt the following interaction design decisions. A PDA with wireless network capability is used and an RFID reader is attached to it to 'scan' the RFID tags used to identify the exhibits. Wi-Fi infrastructure is being used to deliver data and establish communication between the visitors. When an exhibit is scanned, the PDA sends a request for information to the server which delivers the appropriate content presented in the form requested by the user. Data exchange between two users is accomplished through alignment of their devices while pointing one to the other, which mimics the exhibit scanning procedure. We also opted for small chunks of text since reading at low resolution screens reduces reading comprehension significantly.

The educational design of the activity was inspired by the social and cultural perspective of constructivism. It was structured around a set of learning objectives relevant to the thematic focus of the museum, to the exhibits' information, to the age and previous knowledge of the students, and to the fact that involves a school visit (as opposed to individual museum visits). The basic elements which shaped the activity were:

a. **Engagement of interest:** Engagement and interest hold an important role in the learning process. Student interest in a museum should not be taken for granted, especially because a visit arranged by the school is not usually based on the fact that some students might be interested to the theme of the museum. In the inheritance activity we considered to trigger student interest by engaging them in a game. The setting, the rules and the goal of the game were presented in the context of a story.
b. **Building on previous knowledge:** The focus of the activity was selected with respect to the history courses that students were taught in school. They had a general idea about the specific period of the Greek history and the activity offered complementary information about certain issues of this period. Building on previous knowledge was expected to support students in problem solving and hypothesis formulation and testing.
c. **Selecting–processing–combining pieces of information:** The scenario is structured around the idea that students read the offered information, select what is relevant to their inquiry and combine it with other pieces of information that have selected and stored earlier. Thus the students are expected to visit and re-visit the relevant exhibits, go through the information that involves them as many times as they think necessary and not just retrieve that information but combine it and use it in order to find the favorite exhibit which is the end point of the game

d. **Hypothesis formulation and testing:** When students have selected enough information from the exhibits around one room of the museum they can attempt to use some of the clues they have selected in order to find the favorite exhibit. If they fail they can go around the room to collect more information and try again.

e. **Communication and collaboration:** The activity is designed to facilitate inter and intra-group collaboration. Specifically, two groups of students are expected to collaborate to determine which exhibit they will interact with, to exchange clues using their PDA and to discuss their ideas about the favorite exhibit.

During the activity, the participating teams are free to explore any exhibit. Each team is provided with a PDA to extract information related to the exhibits by reading the tags attached to each of the exhibits (Figure 1). Only some of the exhibits contain ‘clues,’ which give information about the favourite exhibit to be found. Children must locate them, store them in the PDAs notepad and after collecting all or most of the clues the teams are able to beam their clues to each other. After collecting all six clues the students are challenged to locate the favorite exhibit. When both teams agree that one exhibit is the favorite one, they can check the correctness of their choice by reading with both PDAs the RFID tag of the chosen exhibit.

After the development of a prototype application, a case study was conducted inside the museum in order to validate the design choices. Seventeen students, aged 11, participated in the study (Figure 2). Data concerning all involving elements were collected to study the activity in depth. The activity was videotaped, PDA screen recording has been used and voice recorders were used to record dialogues among the participants.

The goal of the data analysis was twofold. First, to identify problems children encountered during the process in relation to each of the activity’s elements. Then, to identify the nature of the interactions occurred during the procedure. Our analysis is based on the Activity Theory, concerning mainly human practices from the perspective of consciousness and personal development. It takes into account both individual and collaborative activities, the asymmetrical relation between people and things, and the role of artifacts in everyday life. The activity is seen as a system of human processes where a subject works on an object in order to obtain a desired outcome. In order to accomplish a goal, the subject employs tools, either conceptual or embodiments. Activity is consisted by different components which are (Figure 3): (a) *subject*, (the persons engaged in the activity), (b) *object* (scope of the activity), (c) its *outcome* (c) *tools* used by the subjects (d) *rules-roles* that define the activity process, (e)*community* (context of the activity) and (f)*division of labour* (tasks division among the participants, Kuuti, 1995; Zurita & Nussbaum, 2004).

Figure 1. Screenshots of the “Inheritance” application: (a) Dialogue for RFID tags reading (b) information for a selected exhibit (c) clue selection (d) the notepad screen.

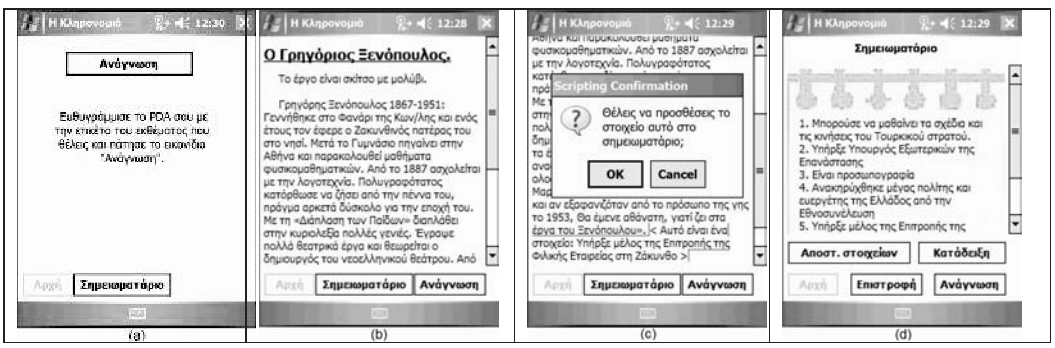


Figure 2. Children engaged in the activity

Activity Theory is of fundamental importance to deeper understand learning with mobile devices while visiting a museum, since in this case knowledge construction is mediated by cultural tools in a social context. The data collected were analyzed with the use of the Collaboration Analysis Tool (ColAT) environment which supports a multilevel description and interpretation of collaborative activities through fusion of multiple data (Avouris, Komis, Fiotakis et al., 2004).

In our analysis, an activity is a procedure during which objects become knowledge through three different levels-steps. Operation is the lower level where routine processes facilitate the completion of goal-oriented actions which in turn constitute the activity. Dialogues, user operations in the application and observations derived from the videos were transcribed in this first level of analysis. The actors, the operations and the mediating tools were noted in this level. Actors were the two participating teams and the researcher. The mediating tools were the dialogue among children and the researcher, texts of information (symbolic) and the application (technological). Some examples of operations in our case are text scrolling, RFID tag reading and transition from one screen to another. Analysis of these user operations led us to the identification of a problem in the use of the application. For example, due to data transfer delay from the server to the PDA, users in some occasions were frustrated and selected repeatedly an action due to lack of timely feedback.

In the second level, the different actions presented among the structural components of the activity are being studied. In order to identify and categorize the actions, a series of typologies were introduced. Typologies were set according to the goal and the mediating tool of each action. For example when children used the PDA to read the text information (mediating tools) their goal was not always the same. Three different typologies were used to describe the situation when the children read carefully the information provided ("Reading of information"), when they were reading the information and searched for clues also ("Reading and searching for clues"), and finally when they were "Searching for clues only." A "Reading and searching for clues" action example is presented in Table 2. Children scroll down the text and one of them states in the end of this action that they were unable to find a clue. When children search only for clues without paying attention to the information we observe rapid scrolling. A clear indication that they have already found all the clues is when children read only the information.

Other actions defined in our study were related to the dialogue between the children and the researcher aimed to overcome difficulties in using the application or understanding the rules-roles of the activity. Typologies where also introduced to describe the interaction between children related to the next step in the procedure (…"*Should we go there? …ok")* and the exchange of thoughts about the solution of the activity (…"*Well, tell me, the first clue is? ...He could spy the Turkish army"...)*. In the third level of analysis, patterns identified concerning the evolution of the procedure.

Figure 3. Description according to the activity theory model

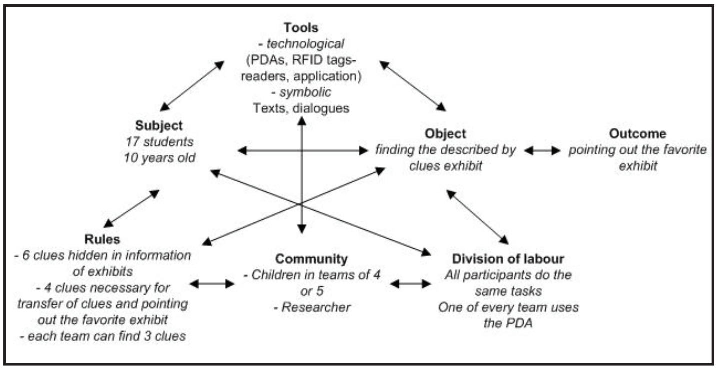


Table 2. An extract of the data analysis presenting action of the 'reading and searching for clues' class

Time	Actor	Tool	Events
00 : 08 : 50	Group1	PDA	Selection of "read"
00 : 08 : 52	Group1	texts	Information D. Stefanou
00 : 09 : 21	Group1	PDA	scrolling
00 : 09 : 33	Group1	PDA	scrolling
00 : 09 : 38	Group1	PDA	scrolling
00 : 09 : 42	Group1	PDA	scrolling
00 : 09 : 45	Group1	PDA	scrolling
00 : 09 : 49	Group1	PDA	scrolling
00 : 09 : 57	Group1	PDA	scrolling
00 : 10 : 02	Group1	PDA	scrolling
00 : 10 : 03	Group1	Dialogue	It doesn't have any (clues) here

Clearly, the basic goals of the activity as described previously in this section have been fulfilled. Data analysis indicated that children were highly motivated by the activity and collaborated in order to achieve their goal. As derived from the analysis, the teams adopted different strategies to accomplish the task. Collaboration was observed mainly while making the choice of the next exhibit to be examined. After completing the task of finding the clues, the two teams collaborated more closely. They divided the work needed to find the exhibit described by the clues and looked in different parts of the room while collaborating and sharing their thoughts and suppositions. They used the clues as information filters thus eliminating the ones that did not match. Additionally, the learning result of this activity, as derived from subsequent students' essays describing the visit experience, was a deeper understanding of the historical role of the persons represented in the exhibits and their interrelations.

CONCLUSION AND FUTURE WORK

This chapter attempted to present current design approaches for mobile learning applications in the context of a museum visit. In addition, thorough study of similar approaches took place, which lead to useful design patterns and guidelines. As discussed, design of mobile learning systems, is not a straightforward process. In addition to the challenge of integrating the concept of context into the design process and independently from context conceptualization, a comprehension of pedagogical goals, desired learning transfers, user typical needs and objectives should take place. We argue that proper design decisions should take into account a solid theoretical cognitive framework, as well as the special characteristics of the mobile devices used and the challenges of such an informal learning setting. A suitable activity should be properly supported by adequate interaction models, deeper understanding of the tasks involved to carry out the activity as a whole and their expectations while carrying out specific actions. For this reason, further validation of our proposed activity, took place in the actual museum. The activity was enjoyed by the students and enhanced their motivation to learn more about the cultural and historical context represented by the exhibits. The latter challenge has been better illustrated while discussing our experience of designing a collaborative learning activity in a cultural history museum and a case study validating its usefulness.

Clearly, the future of learning technology in museums lies in the blending, not the separation, of the virtual and the real world. That is because learning in a museum context could be conceived as the integration, over time, of personal, socio-cultural, and physical contexts. The physical setting of the museum in which learning takes place mediates the personal and socio-cultural setting. The so called ‘interface transparency’ should be treated as an effort to seamlessly integrate the computational device to our natural environment. This goal could be achieved by augmenting physical space with information exchanges, allowing collaboration and communication, enhancing interactivity with the museum exhibits and by seamless integrating instantly available information delivered in various forms. However, the synergy between technology and pedagogy is not straightforward, especially if we take into account the need to tackle issues such as efficient context integration, transparent usage of PDA, and novel pedagogical approaches to exploit the capabilities of the mobile devices. Therefore, further research effort should take place to experience established methods and practices.

REFERENCES

Avouris, N., Komis, V., Fiotakis, G., Dimitracopoulou, A., & Margaritis, M. (2004). Method and Tools for analysis of collaborative problem-solving activities. In *Proceedings of ATIT2004, First International Workshop on Activity Theory Based Practical Methods for IT Design* (pp. 5-16). Retrieved on February 28, 2007 from http://www.daimi.au.dk/publications/PB/574/PB-574.pdf

Chou, L., Lee, C., Lee, M., & Chang, C. (2004). A tour guide system for mobile learning in museums. In J. Roschelle, T.W. Chan, Kinshuk & S. J. H. Yang (Eds.), *Proceedings of 2nd IEEE International Workshop on Wireless and Mobile Technologies in Education—WMTE’04* (pp. 195-196). Washington, DC: IEEE Computer Society.

Ciavarella, C., & Paterno, F. (2004). The design of a handheld, location-aware guide for indoor environments. *Personal Ubiquitous Computing, 8,* 82–91.

Cobb, P. (2002). Reasoning with tools and inscriptions. *The Journal of the Learning Sciences, 11*(2-3), 187-215.

Danesh, A., Inkpen, K.M., Lau, F., Shu, K., & Booth, K.S. (2001). Geney: Designing a collaborative activity for the palm handheld computer. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems—CHI 2001* (pp. 388-395). New York: ACM Press.

Dey, A. (2001). Understanding and using context. *Personal and Ubiquitous Computing Journal, 5*(1), 4-7.

Dillenbourg, P. (Ed.) (1999). *Collaborative learning: Cognitive and computational approaches.* Oxford, UK: Elsevier Science.

Dix, A., Rodden, T., Davies, N., Trevor, J., Friday, A., & Palfreyman, K. (2000). Exploiting space and location as a design framework for interactive mobile systems. *ACM Transactions on Computer-Human Interaction, 7*(3), 285–321.

Dyan, M. (2004). *An Introduction to Art, the Wireless Way*. Retrieved on March 25, 2005 from http://www.cooltown.com/cooltown/mpulse/1002-lasarsegall.asp

Evans, J., & Sterry, P. (1999). Portable computers and interactive multimedia: A new paradigm for interpreting museum collections. *Journal Archives and Museum Informatics, 13*, 113-126.

Feix, A., Göbel, S., & Zumack, R. (2004). Dino-Hunter: Platform for mobile edutainment applications in museums. In S. Göbel, U. Spierling, A. Hoffmann, I. Iurgel, O. Schneider, J. Dechau & A. Feix (Eds.), *Proceedings of the Second International Conference on Technologies for Interactive Digital Storytelling and Entertainment: Conference Proceedings—TIDSE 2004* (pp. 264-269). Berlin: Springer.

Fleck, M., Frid, M., Kindberg, T., Rajani, R., O'Brien-Strain, E., & Spasojevic, M. (2002). From informing to remembering: Deploying a ubiquitous system in an interactive science museum. *Pervasive Computing, 1*(2), 13-21.

Grinter, R. E., Aoki, P. M., Szymanski, M. H., Thornton, J. D., Woodruff, A., & Hurst, A. (2002). Revisiting the visit: understanding how technology can shape the museum visit. In *Proceedings of the 2002 ACM Conference on Computer Supported Cooperative Work—CSCW 2002*, (pp. 146-155). New York: ACM Press.

Hall, T., & Bannon, L. (2006). Designing ubiquitous computing to enhance children's learning in museums. *Journal of Computer Assisted Learning, 22*, 231–243.

Hayhoe, G. F. (2001). From desktop to palmtop: creating usable online documents for wireless and handheld devices. In *Proceedings of the IEEE International Conference on Professional Communication Conference–IPCC 2001* (pp. 1-11).

Klopfer, E., Perry, J., Squire, K., Jan, M., & Steinkuehler, C. (2005). Mystery at the museum: a collaborative game for museum education. In T. Koschmann, T. W. Chan & D. Suthers (Eds.), *Proceedings of the 2005 conference on Computer support for collaborative learning: the next 10 years!* (pp. 316-320). Mahwah, NJ: Lawrence Erlbaum.

Koshizuka, N., & Sakamura, K. (2000). The Tokyo University Museum. In *Kyoto International Conference on Digital Libraries: Research and Practice* (pp. 85-92).

Kuuti, K. (1995). Activity theory as a potential framework for human-computer interaction research. In B. Nardi (Ed.), *Context and consciousness: Activity theory and human computer interaction* (pp. 17-14). Cambridge: MIT Press.

Kwak, S.Y. (2004). *Designing a handheld interactive scavenger hunt game to enhance museum experience*. Unpublished diploma thesis, Michigan State University, Department of Telecommunication, Information Studies and Media.

Laurillau, Y., & Paternò, F. (2004). Supporting museum co-visits using mobile devices. In S. Brewster & M. Dunlop (Eds), *Proceedings of the 6th International Symposium on Mobile Human-Computer Interaction—Mobile HCI 2004* (pp 451-455). Berlin: Springer.

Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. New York: Cambridge University Press.

Luchini, K., Quintana, C., Krajcik, J., Farah, C., Nandihalli, N., Reese, et al. (2002). Scaffolding in the small: Designing educational supports for concept mapping on handheld computers. In *CHI 2002 Extended Abstracts on Human Factors in Computing Systems* (pp. 792-793). New York: ACM Press.

Luyten, K., & Coninx, K. (2004). ImogI: Take control over a context aware electronic mobile guide for museums. *In proceedings of the 3rd Workshop on HCI in Mobile Guides*. Retrieved on February 24, 2007 from http://research.edm.luc.ac.be/~imogi/

Myers, B. A., Stiel, H., & Gargiulo, R. (1998). Collaboration using multiple PDAs connected to a PC. In *Proceedings of the ACM 1998 Conference on*

Computer Supported Cooperative Work — CSCW '98 (pp. 285-294). New York: ACM.

Mase, K., Sumi, Y., & Kadobayashi, R. (2000). The weaved reality: What context-aware interface agents bring about. In *Proceedings of the Fourth Asian Conference on Computer Vision - ACCV2000* (pp. 1120-1124).

Norman, D. A. (1986). Cognitive engineering. In D. A. Norman & S. W. Draper (Eds.), *User centered systems design* (pp. 31-61). Mahwah, NJ: Lawrence Erlbaum.

Norris, C., & Soloway, E. (2004). Envisioning the handheld-centric classroom. *Journal of Educational Computing Research, 30*(4), 281-294.

Oppermann, R., Specht, M., & Jaceniak, I. (1999). Hippie: A nomadic information system. In H. W. Gellersen (Ed.), *Proceedings of the First International Symposium Handheld and Ubiquitous Computing - HUC'99* (pp. 330-333). Berlin: Springer.

Patten, B., Arnedillo Sanchez, I., & Tangney, B. (2006). Designing collaborative, constructionist and contextual applications for handheld devices. *Computers and Education, 46*, 294-308.

Perry, D. (2003). *Handheld Computers (PDAs) in Schools.* BECTA ICT Research Report. Retrieved on February 26, 2007 from http://www.becta.org.uk/ page_documents/research/handhelds.pdf

Raptis, D., Tselios, N., & Avouris, N. (2005). Context-based design of mobile applications for museums: a survey of existing practices. In M. Tscheligi, R Bernhaupt & K. Mihalic (Eds.), *Proceedings of the 7th international Conference on Human Computer interaction with Mobile Devices & Services- Mobile HCI 2005* (pp. 153-160). New York: ACM Press.

Rieger, R., & Gay, G. (1997). Using mobile computing to enhance field study. In R.P. Hall, N. Miyake & N. Enyedy (Eds.), *Proceedings of Computer Support for Collaborative Learning –CSCL 1997* (pp. 215–223). Mahwah, NJ: Lawrence Erlbaum.

Rocchi, C., Stock, O., Zancanaro, M., Kruppa, M., & Krüger, A. (2004). The museum visit: Generating seamless personalized presentations on multiple devices. In J. Vanderdonckt, N. J. Nunes & C. Rich (Eds.), *Proceedings of the Intelligent User Interfaces - IUI 2004* (pp. 316-318). New York: ACM.

Roschelle, J. (2003). Unlocking the learning value of wireless mobile devices. *Journal of Computer Assisted Learning, 19*(3), 260-272.

Thom-Santelli, J., Toma, C., Boehner, K., & Gay, G. (2005). Beyond just the facts: Museum detective guides. In *Proceedings from the International Workshop on Re-Thinking Technology in Museums: Towards a New Understanding of People's Experience in Museums* (pp. 99-107). Retrieved on February 25, 2007 from http://www.idc.ul.ie/museumworkshop/programme.html

Vahey, P., & Crawford, V. (2002). *Palm education pioneers program final evaluation report*. Menlo Park, CA: SRI International.

Van Gool, L., Tuytelaars, T., & Pollefeys, M. (1999). Adventurous tourism for couch potatoes. (Invited). In F. Solina & A. Leonardis (Eds.), *Proceedings of the 8th International Conference on Computer Analysis of Images and Patterns – CAIP 1999* (pp. 98-107). Berlin: Springer.

Yatani, K., Sugimoto, M., & Kusunoki, F. (2004). Musex: A System for Supporting Children's Collaborative Learning in a Museum with PDAs. In J. Roschelle, T.W. Chan, Kinshuk & S. J. H. Yang (Eds.), *Proceedings of 2nd IEEE International Workshop on Wireless and Mobile Technologies in Education – WMTE'04* (pp 109-113). Washington, DC: IEEE Computer Society.

Zurita, G., & Nussbaum, M. (2004). Computer supported collaborative learning using wirelessly interconnected handheld computers. *Computers and Education, 42*, 289-314.

KEY TERMS

Activity Theory: Is a psychological framework, with its roots in Vygotsky's cultural-historical psychology. Its goal is to explain the mental capabilities of a single human being. However, it rejects the isolated human being as an adequate unit of analysis, focusing instead on cultural and technical mediation of human activity.

Context: Context is any information that can be used to characterize the situation of an entity. An entity is a person, place, or object that is considered relevant to the interaction between a user and an application, including the user and applications themselves (Dey, 2001).

Context-Aware: The ability to sense context.

Interaction Design: Interaction design is a sub-discipline of the design notion which aims to examine the role of embedded behaviors and intelligence in physical and virtual spaces as well as the convergence of physical and digital products. In particular, interaction design is concerned with a user experience flow through time and is typically informed by user research design with an emphasis on behavior as well as form. Interaction design is evaluated in terms of functionality, usability and emotional factors.

Mobile Device: A device which is typically characterized by mobility, small form factor and communication functionality and focuses on handling a particular type of information and related tasks. Typical devices could be a Smartphone or a PDA. Mobile devices may overlap in definition or are sometimes referred to as information appliances, wireless devices, handhelds or handheld devices.

Mobile Learning: Is the delivery of learning to students who are not keeping a fixed location or through the use of mobile or portable technology.

Museum Learning: A kind of informal learning which is not teacher mediated. It refers to how well a visit inspires and stimulates people into wanting to know more, as well as changing how they see themselves and their world both as an individual and as part of a community. It is a wide concept that can include not only the design and implementation of special events and teaching sessions, but also the planning and production of exhibitions and any other activity of the museum which can play an educational role.

Chapter XVII
Collaborative Learning in a Mobile Technology Supported Classroom

Siu Cheung Kong
The Hong Kong Institute of Education, Hong Kong

ABSTRACT

This chapter introduces the migration of a Web-based cognitive tool (CT) for the generation of procedural knowledge about mathematical fractions from a desktop version to a mobile version. It aims to provide insight into the potential of human-computer interaction in mobile learning environments to encourage reciprocal tutoring and foster collaborative learning. A collaborative mobile learning environment is designed using a design-based research approach. A Web-based CT for learning the concept of fraction equivalence is improved and modified to suit the environment as applied to a mobile technology supported classroom. This chapter first delineates the theoretical design approach and empirical design methodology that underlie the migration exercise, and then discusses the architectural design of artifacts and the pedagogical design of learning activities to shed light on the development and application of mobile technology in a classroom learning environment.

INTRODUCTION

Procedural knowledge is the knowledge that guides the performance of a task in the absence of access to the knowledge that underlies the procedure (Anderson, 1976). To acquire procedural knowledge about the operation of mathematical fractions, it is necessary to first have knowledge of fraction equivalence, which comprises the concept of fraction equivalence and knowledge of the computation of equivalent fractions, both of which are of equal importance (Kong & Kwok, 2005). Procedural knowledge of adding fractions with unlike denominators is more likely to be generated if a conceptual understanding of fraction equivalence is initially developed (Kong & Kwok, 2005).

Early research shows that learners seldom easily understand the procedural knowledge that is associated with fraction operations, such as addition and subtraction (Huinker, 1998; Niemi, 1996; Pitkethly & Hunting, 1996). Traditional classroom instruction in this topic generally adopts the algorithmic approach, which suffers from the shortcoming of separating knowledge from meaning. To rectify this problem, a Web-based cognitive tool (CT) was developed to assist learners to generate procedural knowledge of adding fractions with unlike denominators (Kong, 2001) with the rationale that CTs are both mental and computational devices that can support, guide, and mediate the cognitive processes of learners (Derry & Lajoie, 1993; Kommers, Jonassen & Mayes, 1992).

Previous evaluation studies (Kong & Kwok, 2002, 2005) show that the adoption of reciprocal tutoring in a collaborative learning environment has the potential to increase learning effectiveness in this domain. As the portable nature of mobile devices offers the opportunity to promote reciprocal tutoring in a mobile technology supported classroom, the desktop version of the Web-based CT for comprehending procedural knowledge of mathematical fractions is adapted to create a mobile version for collaborative learning.

DESIGN FRAMEWORK

The goal of cognitive technology is to develop CTs that meet the needs of human users (Janney, 1999; Pea, 1985). The capability of the aforementioned Web-based CT to assist learners to generate procedural knowledge of adding fractions with unlike denominators has been validated, and experimental studies have revealed that it serves as a mediator that triggers discussion among learners. Slavin (1996) states that a collaborative learning context, such as discussion, is an important way of stimulating reflection among learners. From the perspective of cognitive science, peer discussion is a way of stimulating cognitive elaboration (Wittrock, 1979). In light of these views, the aim of the study presented in this chapter is to further improve the CT to meet the needs of learners who are learning naturally in a classroom setting by applying the tool to a collaborative learning environment in a mobile technology supported classroom. This section outlines the design framework of the new CT, which, when used in a mobile learning environment, promotes collaborative engagement and encourages the resolution of cognitive conflict by cognitive elaboration and reciprocal tutoring in the classroom.

Cognitive Elaboration

Cognitive elaboration is the process of forming associations between new information and prior knowledge. It is regarded as an essential process for facilitating comprehension and knowledge acquisition (Wittrock, 1986). In cognitive models, learners play an active, responsible, and accountable role in their generative learning. Newly learned materials are better retained and more easily recalled if learners undergo spontaneous cognitive elaboration to trace the relations between the new information and known information, because cognitive elaboration helps the transfer of new information from the short-term to the long-term memory (Doherty, Hilberg, Pinal et al., 2003; Wittrock, 1979).

An effective means of fostering the capability of learners to cognitively elaborate is to offer them opportunities to practice cognitive elaboration. Reciprocal tutoring is a good strategy for encouraging learners to practice cognitive elaboration in which learners take turns to tutor each other in a group learning context (Chan & Chou, 1997; Wong, Chan et al., 2003). The strategy enables students to learn from one another through the verbal elaboration of the new knowledge in a group learning context, thus allowing students who have gained insight into the new concept to reinforce their knowledge by providing explanations to others who need more opportunity to comprehend the knowledge. In this way, all group members benefit from engaging in the elaboration process.

Cognitive Conflict

Understanding mathematical ideas often involves the restructuring of the mathematical schema of learners. This restructuring process is intricately linked with the occurrence of cognitive conflict (Tall, 1977). Cognitive conflict is a tenet of the psychological theory of cognitive change, and is an inferred state of incompatibility between two inferred component states within the cognitive process (Cantor, 1983). In general, cognitive conflict is a perceptual state in which one notices the discrepancy between an anomalous situation and a preconception (Lee, Kwon, Park, et al., 2003). Since the 1980s, the exploitation of cognitive conflict has been regarded as a feasible teaching strategy because the cognitive change that such conflict creates induces introspection among learners about the newly learned conception and its incongruity with their preconceptions or misconceptions.

According to the cognitive conflict process model, there are three stages in the engagement of cognitive conflict in learning (Lee et al., 2003). The first stage is the preliminary stage, where an anomalous situation that differs from the preconceptions of learners is introduced. The second stage is the conflict stage, in which learners recognize and reappraise the anomalous situation and express anxiety or interest in resolving the conflict. The third stage is the resolution stage, in which learners try to resolve cognitive conflict in any way that they can.

Cognitive conflict has constructive, destructive, and meaningless potential. When learners clearly recognize an anomaly and reappraise a situation of cognitive conflict deeply by expressing strong interest or anxiety, the cognitive conflict has constructive potential. When learners do not recognize the anomaly or simply ignore it and express feelings of frustration or rejection, the cognitive conflict is regarded as destructive. When learners recognize the anomaly, but accept it passively without interest or cognitive reappraisal, the cognitive conflict is regarded as meaningless. Early studies show that the inducement of constructive cognitive conflict promotes positive outcomes in classroom learning (Limón, 2001). However, the creation of constructive cognitive conflict largely depends on the interdependence of learners, and is thus closely related to collaborative learning processes that provide learners with ample opportunity to learn from peer discussion.

From the perspective of cognitive science, cognitive elaboration has a positive effect on cognitive conflict. In cognitive models, individual differences between learners are important, especially when they involve cognitive processes (Wittrock, 1979), as they can be a source of cognitive conflict. In a discussion during which cognitive conflict occurs, learners learn from one another through peer interaction, and are provoked to reflect through argument. Understanding then emerges through mutual elaboration of the new concept. The process of cognitive elaboration therefore induces cognitive conflict with a positive potential.

As the construction of knowledge of fraction equivalence is often accompanied by the occurrence of cognitive conflict (Kong & Kwok, 2002), this study aims to design a collaborative learning environment to help learners to generate procedural knowledge of adding fractions with unlike denominators in this learning context.

Collaborative Learning

Collaborative learning is a process that encourages learners to participate in coordinated and synchronous learning activities with a number of other learners (Roschelle & Teasley, 1995). It emphasizes the concept that "every learner learns from everyone else" (Fischer, Bruhn, Gräsel et al., 2002, p. 215), and promotes self-directed and active learning through group learning activities that require interdependence among group members. The basic assumption behind collaborative learning is that learners are ready to interact with one another to offer help or share ideas. According to this rationale, learners value and encourage group members in the learning process, and thus depend on one another to achieve effective learning through collaborative interaction with their peers (Johnson & Johnson, 1999; Slavin, 1996).

There are several characteristics of collaborative learning, such as sharing knowledge among peers, mediation by teachers, and the arrangement of learners into heterogeneous groups (Dillenbourg, 1999). In collaborative learning, learners take the role of knowledge provider by sharing their own knowledge and learning strategies with other group members, and teachers play the role of facilitator by providing mediation for group learning, such as adjusting the information flow or level of interaction among groups and group members. The heterogeneous grouping of learners is important in collaborative learning, because it allows reciprocal tutoring and knowledge exchange, which helps learners to develop knowledge and interpersonal communication skills. Although collaborative learning has been found to be effective for learners at all learning achievement levels (Slavin, 1996), the two obstacles of problems with class control during active participation and the unsatisfactory participation of particularly quiet learners can decrease the effectiveness of the learning process (Roschelle, 2003).

Mobile Learning

Mobile learning refers to the use of mobile technology for learning and teaching. It is an emergent learning approach that has the potential to address the aforementioned two obstacles to collaborative learning. Portability and versatility make mobile devices a powerful medium for learning (Sharples, Taylor & Vavoula, 2005), and a number of design-based studies have shown the prospect of mobile learning in education (Roschelle, 2003). Mobile technology has two attributes that facilitate the design of collaborative learning activities in a classroom environment.

First, the portability of mobile devices offers learners a sense of ownership of individual mobile devices. Mobile learning enables learners to hold the mobile devices during the entire lesson in the classroom. This sense of ownership helps to provide incentives to learners to actively participate in collaborative learning activities. In addition, with the use of non-verbal communication features, such as graphical support in CT, of the mobile devices at various stages of the learning process, the use of mobile devices may trigger all learners, including the quiet ones, to participate in the learning process. This increases the opportunities for the inducement of cognitive conflict and cognitive elaboration.

Second, the versatility of the currently-available mobile devices facilitates collaborative learning activities in a coordinated manner. This can be achieved by creating continuous learning process to cater for learning diversity, and providing flexibility for teachers to rearrange the groupings of learners. By programming the mobile technology supported collaborative learning environment, learners may continually receive questions generated with appropriate difficulty levels and constantly receive judgment on the correctness of learners' responses, teachers are thus prevented from the chaotic class order caused by learning diversity of learners in the active learning process. With appropriate programming effort, such as data mining of collaborative learning, teachers may obtain information from the system to regroup learners into heterogeneous groups. These supports allow teachers to have more time to manage the learning progress of learners.

DESIGN METHODOLOGY

This study adopts a design-based research approach to designing the collaborative mobile learning environment. Design-based research is a fundamental mode of scholarly inquiry that is useful in many academic fields that in the past decade has become an increasingly accepted approach to theoretical and empirical study in the field of education (Bell, 2004). Design-based research is an attempt to combine empirical educational research with the theory-driven design of learning environments (Bell, 2004; Hoadley, 2004; Design-Based Research Collective, 2003) to design and explore a whole range of innovations. The key components of design-based research include architectural design, such as the artifacts that are involved in learning activities, and pedagogical design, such as the structure and scaffolding of learning activities (Design-Based Research Collective, 2003).

The study presented in this chapter is motivated by the results of empirical educational research on the effectiveness of a CT for learning the concepts of mathematical fractions. A qualitative pilot evaluation study and a quasi-experimental evaluation study that investigated how learners develop a concept of mathematical fractions and acquire procedural knowledge of fraction operations by using a CT named the "Graphical Partitioning Model" (GPM) form the design basis of the mobile CT.

The CT for supporting the generation of procedural knowledge about fraction addition that is introduced in this chapter originates from a cognitive task analysis of the domain (Kong, 2001) that initially led to the design of the GPM CT, which is a rectangular bar with partitioning capability.

The pedagogical benefit of the GPM is that it reveals the procedural structure for evaluating fraction expressions by linking the concrete manipulation of further partitioning the fraction bars to find a common fractional unit with the meaning of finding a common denominator in the process of adding fractions with unlike denominators. To create descriptive meanings and features that enable the learning effectiveness of the CT to be evaluated, a qualitative pilot evaluation study was conducted (Kong & Kwok, 2002). The evaluation study aimed to investigate how learners use the GPM to develop a concept of fractions and to understand and acquire procedural knowledge of fraction operations.

In the pilot evaluation study, 12 subjects with various levels of mathematical ability were selected to use the CT to learn about the subject domain for five successive two-hour sessions. Learners used the CT to acquire knowledge of fraction equivalence in the first three sessions of the course, and to develop knowledge of adding fractions with unlike denominators in the last two sessions. Learners used the partitioning capability of the CT to explore the concept of fraction equivalence and the procedural knowledge of adding fractions.

The results of the study indicate that the GPM only benefited learners who were already able. These learners gained knowledge of fraction equivalence by working with the CT, and acquired procedural knowledge about adding fractions with unlike denominators. The investigation of the knowledge profile of the learners in developing the concept of fraction equivalence (including cognitive artifacts such as diagrams that were drawn by the learners in the attainment tests, worksheets, and post-test performance results) and observation of the initiative of learners in using equivalent fractions to add fractions with unlike denominators yielded three key findings. First, half of the learners did not understand the inverse relationship between the number of parts and the size of a part of a unit. Second, almost 60 percent of the learners showed no intention of representing fractions to compare their equivalence. Third, around 70 percent of the learners showed no ability to use equivalent fractions to add fractions with unlike denominators, although some understood the concept of fraction equivalence but were unable to relate the concept to ways of finding equivalent fractions. These findings reveal that to improve the effectiveness of the CT in helping learners to learn the procedural knowledge of adding fractions with unlike denominators, its capability to teach fraction equivalence must be enhanced.

In response to the three key findings of the pilot evaluation study, the Web-based CT was improved to make it into a model of affordances to support the learning of mathematical fractions. This enhanced version of the Web-based CT forms the basis of the mobile version of the CT that is introduced in this chapter. From the constructivist perspective, the means of instruction should not be predetermined, because each learner constructs knowledge in a unique way, and thus "the pedagogical role of the system is to provide profitable spaces for interaction to the learner based on some model of the affordances of potential situations" (Akhras & Self, 2000, p. 24). Gibson (1979) introduces the notion of affordances and suggests that the perceptual task of human beings is to detect environmental aids that could be used in their attempts to interact with the environment to meet their needs. The function of a model of affordances is to make available profitable spaces, or provide the necessary scaffolding (Clark, 1997). Therefore, to enhance the effectiveness of the CT to support the learning of mathematical fractions, a model of affordances for the teaching of fraction equivalence to help develop procedural knowledge about adding fractions with unlike denominators was designed to give learners the means to interact in a way that meets their needs.

The enhanced CT underwent three modifications to improve its ability to stimulate knowledge of fraction equivalence in each learner according to that learner's particular needs (Kong & Kwok, 2003). The first modification was the addition of a space for partitioning that allows a choice to be made between an intentional slowed down animation that shows the partitioning or regrouping process and an instantaneous change that shows the results of the partitioning or regrouping process. The simulation of the partitioning strategy by the slowed animation addresses the lack of intention of representing fractions to compare their equivalence and the failure to recognize the inverse relationship between number of parts and the size of a part of a unit that were identified in the evaluation study. This modification allows learners to interact with the CT according to their needs, with capable learners being able to generalize the knowledge by rapidly calling up the results of partitioning and less capable learners being able to pick up the idea by activating the slowed down animation of the process of partitioning.

The second modification was the addition of a space for the comparison of the equivalence of fractions in response to the difficulties that learners who have no intention of representing fractions with the same unit to compare their equivalence encounter. An animation that shows the direct comparison of the equivalence of two fraction bars, which is triggered by dragging a fraction bar and dropping it onto another bar, was designed to allow an extra comparison of fraction equivalence in addition to the visual inspection of two separate fraction bars. This modification gives learners multiple opportunities to compare fraction equivalence in an interactive way.

The third modification was the addition of a space that consists of a hypothesis-testing interface $\frac{a}{b} = \frac{a \times c}{b \times d}$ to address the problem of learners who lack the ability to find equivalent fractions systematically. The hypothesis-testing interface asks learners to test possible fraction equivalent states by adjusting parameters *c* and *d*, and allows them to compare fraction equivalence using the aforementioned comparison animation by dragging the fraction bar of the hypothesized fraction and dropping it onto the fraction bar of the original fraction.

The enhanced CT was evaluated by a quasi-experimental evaluation study with a pre-test/post-test control group design (Kong & Kwok, 2005). The results of this evaluation study indicate that the model of affordances allows learners of varying learning abilities to develop a concept of fraction equivalence. It was observed that with the mediation of the enhanced CT, learners were able to generate procedural knowledge about adding fractions with unlike denominators.

Using the empirical findings on the effectiveness of the model of affordances for teaching fraction equivalence and the theoretical background that is discussed in the previous sections, in the following sections we discuss the architectural and pedagogical design of a collaborative learning environment in a mobile technology supported classroom.

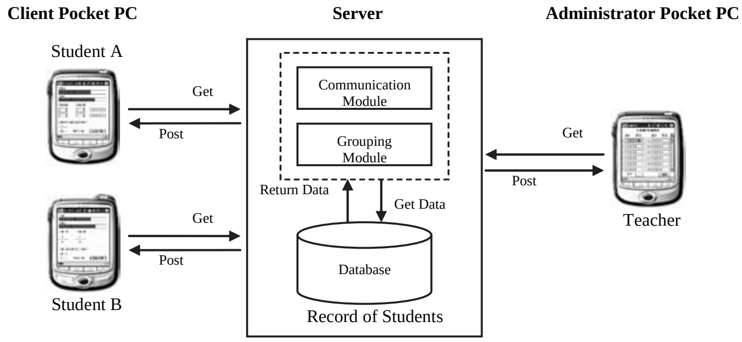
DESIGN OF A COLLABORATIVE LEARNING ENVIRONMENT FOR COGNITIVE ENGAGEMENT

A Mobile Technology Supported Classroom

Collaboration is a coordinated, synchronous activity in which continued attempts are made to construct and maintain a shared conception of a problem (Roschelle & Teasley, 1995). In the study introduced in this chapter, a series of synchronous interactions in a mobile technology supported classroom is designed to encourage learners to engage in learning tasks, and a mobile platform is established that allows immediate interaction between learners working in pairs. This pair-wise grouping aims to induce in-depth discussion between group members. Figure 1 depicts the mobile technology supported classroom.

The learning activities take place in a wireless-networked classroom, and the mobile device that is used is a pocket PC, which is chosen for its

Figure 1. Mobile technology supported classroom for learning fraction equivalence in pairs



portability and relatively large screen. The teacher and learners are provided with pocket PCs. The teacher's pocket PC is installed with an interface for managing the pairs and for organizing the learning activities. The pocket PCs that are used by the learners are installed with a graphical tool for learning fraction equivalence. The learners interact in pairs through a server that is connected to an SQL database. The server coordinates the grouping instructions of the teacher and the synchronous interactions between paired learners.

A Model of Affordances for Teaching Fraction Equivalence

Knowledge of fraction equivalence is a prerequisite for the development of procedural knowledge of the addition of fractions with unlike denominators. Therefore, collaborative learning activities with graphical support are designed to develop the understanding of learners in this domain. In the learning activities, a graphical model of a rectangular bar is used to represent fractions (Kong & Kwok, 2003), with each fraction being represented by displaying shaded fractional parts of an equally partitioned rectangular bar according to the value of the fraction. The three spaces of the model of affordances that will be discussed are adapted from the desktop version of the Web-based CT.

First Space: Learner-Controlled Animation of Partitioning

The cognitive artifact of the first space is an animation that is adapted from the graphical partitioning capability of the Web-based CT and is designed to address the lack of understanding among learners of the part-whole concept and the inverse relationship between the number of parts in a unit and the size of a part. Figures 2a to 2d demonstrate this feature. The learner-controlled animation allows learners to partition fractions by clicking the graphical representation of the fractions. When learners click on the graphical representation (see Figure 2a), the fraction bar becomes blank to show the initial state of the fraction as a unit (see Figure 2b). When learners click on the blank rectangular bar, the bar is partitioned into fractional units (see Figure 2c). This allows learners to develop a concept of fractions as parts of a whole. When learners click on the calibrated fraction bar, the parts of a unit are shaded based on the value of the numerator (see Figure 2d). This feature helps learners build up the concept that the parts of the whole are equal, and to understand the inverse relationship between the number of parts and the size of a part of a unit, that is, that the larger the denominator, the smaller the size of a part in the unit.

Figure 2.

Figure 2a: Graphical representations of the display of two fractions on the interface.

Figure 2b: A blank rectangular bar is shown when a learner clicks on a fraction bar.

Figure 2c: A calibrated fraction bar is shown when a learner clicks on the blank fraction bar.

Figure 2d: The original fraction bar with shaded fractional part is shown when a learner clicks on the calibrated fraction bar.

This cognitive artifact is incorporated in all of the fraction bars that appear in the learning activities that involve graphical support. When learners work with only one fraction bar, the stepwise design helps them to develop the part-whole concept, and when they work with both fraction bars, the stepwise design helps them to understand the inverse relationship between the number of parts in a unit and the size of a part. This design is better than that of the desktop version in that it returns control of the learning process to the learners.

Second Space: Comparison of the Equivalence of Fractions

The mobile tool has two cognitive artifacts that are designed to address the difficulties that learners have in comparing the equivalence of fractions. The first is an animation that compares the equivalence of fractions that are represented in visual form, and is illustrated in Figure 3.

Button "A" in the top-right corner starts the comparison animation. Learners can press this button in order to obtain a graphical representation of a fraction, the bar of the selected fraction rolls over the other fraction bar to allow learners to compare the equivalence of the two fractions visually. This feature differs from the desktop version only in the positioning of the two fraction bars for comparison, in that the fraction bars in the desktop version are arranged in a row, whereas those in the mobile version are arranged in a column because of the relatively narrower screen of the mobile device.

The second artifact is the random display of fraction bars with non-comparable graphical representations that is given when learners are asked to compare fractions in the learning activities. An additional feature of an adjustable fraction bar is included to strengthen the concept of representing fractions using common units to compare their equivalence. Figures 4a to 4c demonstrate the adjustable fraction bar for representing fractions.

To compare the equivalence of two fractions, learners must represent fractions using the same unit. An adjustable bar that is 50 percent to 70 percent of the length of the bar of the other fraction is displayed. Button "B" in the top-right corner of the adjustable fraction bar adjusts the length of the bar, and elongates it to the length of the other fraction bar (see Figure 4b) when clicked by a learner. If the learner clicks button "B" again, then the adjustable fraction bar is shortened back to its original length (see Figure 4c). This feature acts as a random alert to assist learners to develop the awareness of the prerequisite of comparing fraction equivalence. There is a similar feature in the desktop version of the CT.

Figure 3. Animation for comparing the equivalence of two fractions

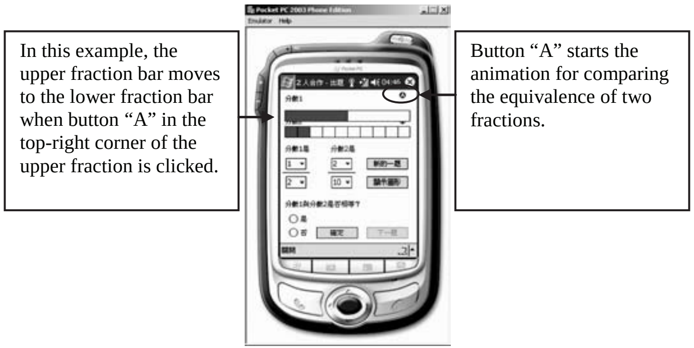


Figure 4.

Figure 4a: The lower fraction bar in this example is the adjustable fraction bar. Its length is 50% to 70% that of the upper fraction bar.

Figure 4b: The adjustable fraction bar elongates to the length of the upper fraction bar when button "B" in the top-right corner of the adjustable fraction bar is clicked.

Figure 4c: The lengthened adjustable fraction bar reduces to its original length when button "B" in the top-right corner of the adjustable fraction bar is clicked again.

Third Space: Time-Keeping Hypothesis-Testing Competition

The third space of the model of affordances comprises a hypothesis-testing interface for finding equivalent fractions. This cognitive artifact is designed to assist learners to develop the capability to compute equivalent fractions. Knowledge of fractions, as aforementioned, consists of two parts: the concept of fraction equivalence and knowledge of the computation of equivalent fractions. To develop knowledge of the computation of equivalent fractions, a time-keeping learning activity is designed in which each pair of learners is engaged in a competition to find an equivalent fraction of a fraction that is assigned by the computer system.

This learning activity involves two steps. Step 1 is the process of finding the equivalent fraction. In this step, learners require to find an equivalent fraction with the use of a hypothesis-testing interface $\frac{a}{b} = \frac{a \times c}{b \times d}$. Learners are asked to adjust parameters *c* and *d* to test a possible fraction equivalent state. Graphical representations are generated to help learners to compare the equivalence of the two fractions by the instant change that takes place following the adjustment of parameters *c* and *d*. Once learners have decided on their answer, then they can click the "Confirm" button to send the answer to the server.

Step 2 is the process of judgment. In this step, the computer system measures the response time and judges the correctness of the answers that are provided by learners. For quick responses that are correct the words "Correct" and "Yeah!" are displayed on the screen (see Figure 5a); for slow responses that are correct, the words "Correct" and "Cheer up!" are displayed (see Figure 5b); and for incorrect answers the words "Incorrect" and "Cheer up!" (see Figure 5c) are displayed regardless of the response time. This learning activity offers learners the opportunity to engage in cognitive elaboration to find out the algorithm for the computation of equivalent fractions, and the use of peer competition helps learners to relate the concept of fraction equivalence to the algorithm for the computation of equivalent fractions. There is a similar interface in the desktop version of the CT, but the interface in the mobile version is improved by the use of competition between learners to encourage cognitive elaboration.

Pedagogical Design: Two Situations for Reflection and Cognitive Elaboration

In this section, we discuss how the pedagogical design of the collaborative learning environment stimulates cognitive conflict and cognitive elaboration. In a learning activity that involves understanding the concept of fraction equivalence, learners are placed in a situation in which they have to decide the equivalence of two fractions. The learners are arranged in pairs: one learner is the question-setter and the other the respondent. The learners alternate between the two roles.

The learning activity comprises three steps. Step 1 is the process of question-setting, in which the learners who are playing the role of question-setter set and send out questions about the equivalence of two fraction expressions. The question-setters have to state whether the two fraction expressions that they have chosen are equivalent by graphically representing the two fraction expressions at the top of the interface. Once the learners are satisfied with the question that they have set, they can click the "Confirm" button to send the question to their partners through the server. Step 2 is the response process. In this step, the learners who are playing the role of respondent receive the question in the form of two fraction expressions from their partner, and then have to decide whether the two expressions are equivalent with the help of visual representation. After indicating their decision, learners click the "Confirm" button to send their answer to the server. Step 3 is the process of judgment. In this step, the computer system assesses the correctness of the questions that are set by the question-setters and the answers that are provided by the respondents. The computer system then sends the messages in the form of the words "Correct" and "Incorrect" for right and wrong questions or answers, respectively.

Figure 5.

Figure 5a: The words "Correct" and "Yeah!" are displayed for correct and quick responses.

Figure 5b: The words "Correct" and "Cheer up!" are displayed for correct but slow responses.

Figure 5c: The words "Incorrect" and "Cheer up!" are displayed for incorrect response regardless of the response time.

Figure 6.

Figure 6a: The computer system generates the message “Please Discuss” if group members have different views.

Figure 6b: Each learner must click the “Discussion Finished” button after the discussion has ended.

Two types of cognitive conflict may be engendered by the learning activity: one is triggered by anomalies between learning peers and the other is triggered by anomalies between learners and the computer system. When one member of a pair of learners provides the correct question or answer and the other gives the wrong question or answer, then the computer system displays the message “Please Discuss” (see Figure 6a). This generates the first type of cognitive conflict and invites learners to share their understanding and to engage in self-reflection and negotiation through collaborative interaction.

When both members agree that they have finished their discussion, they click the “Discussion Finished” button (see Figure 6b) to inform the computer system. The computer system then tells the question-setter and respondent whether they are “Correct” or “Incorrect” (see Figures 7a and 7b). When they differ from the judgments of the learners, these authority judgments create the second type of cognitive conflict, which offers learners a second opportunity to engage in self-reflection and to share their understanding through a post-task discussion.

The aim of the learning activity is to equip learners with a basic knowledge of fraction equivalence through collaborative learning in a mobile technology supported classroom. The activity emphasizes the sharing of knowledge among learners using graphical support to aid the learning process. The different rates of progress of individual learners determine the learning progress of each group. Figure 8 depicts the different statuses of groups of learners in the process of learning the concept of fraction equivalence.

Case 1 is expected to occur commonly at the beginning of the learning process. In this case, both members of the group have a preconception or misconception about the equivalence of fractions, and always set and reply to questions incorrectly. In this situation, both learners encounter the first and second type of cognitive conflict, and such groups can be categorized as being at learning status 1. Cases 2 and 3 occur when one of the group members begins to grasp the concept of fraction equivalence better than his or her partner. The learner who has understood the concept of fraction equivalence begins to set correct questions and give correct answers, whereas his or her counterpart cannot always achieve this status. Both learners in this situation experience the first type of cognitive conflict, and in addition the learner who has yet to understand the concept of fraction equivalence experiences the second type of cognitive conflict, in which an anomalous situation exists between the learner and the judgment of the computer system. Such

Figure 7.

Figure 7a: The computer system generates the message "Correct" if a right question is set.

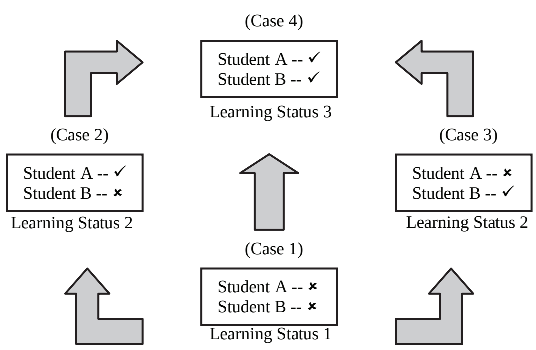
Figure 7b: The computer system generates the message "Incorrect" for a wrong answer.

groups can be categorized as being at learning status 2. Case 4 occurs when both learners in the group have a good understanding of the concept of fraction equivalence and always set and reply to questions correctly. In this case, cognitive conflict rarely occurs, and groups in this situation achieve learning status 3, which is the learning goal of all of the groups. Some groups may go through learning status 1 and 2 to reach learning status 3, and some may go directly from learning status 1 to status 3. The groups in learning status 2 are heterogeneous groupings in this study.

Pedagogical Design: Encouraging Reciprocal Tutoring

The goal of the learning activities that are designed to teach fraction equivalence is to help all of the groups of learners to attain learning status 3 through a collaborative learning environment. The groups in learning status 1 and 2 require the attention and mediation of teachers to promote productive knowledge sharing. In this collaborative learning environment, teachers play the role of mediator, rather than the authority that judges

Figure 8. Status of groups of learners in the process of learning the concept of fraction equivalence



the correctness of answers, and serve to encourage reciprocal tutoring, which can be achieved more productively in a heterogeneous group context and by the promotion of tutoring activities. Reciprocal tutoring enhances the quality of arguments between group members and results in the inducement of more constructive cognitive conflict. Two pedagogical tools are therefore designed to encourage the reciprocal tutoring of learners. The first is the re-grouping of learners and the second is the alteration of the question-setting mode. Figure 9 shows the teacher's interface for the re-grouping of learners and the changing of the question-setting mode.

If teachers observe groups that are stuck at learning status 1 (in which both group members are struggling with a concept), then they can use the first pedagogical tool to swap one of the group members with a member from a group at learning status 3. This helps to achieve more heterogeneous groups, which in turn helps to encourage prolific reciprocal tutoring.

For groups at learning status 2 (in which one of the group members consistently designs incorrect questions or provides incorrect answers), teachers can use the second pedagogical tool to designate another learner as the sole question-setter by changing the mode of question-setting from "Turn-Taking" to "Designation." This creates an environment that allows learners with a better understanding to tutor learners who are still developing the required concept.

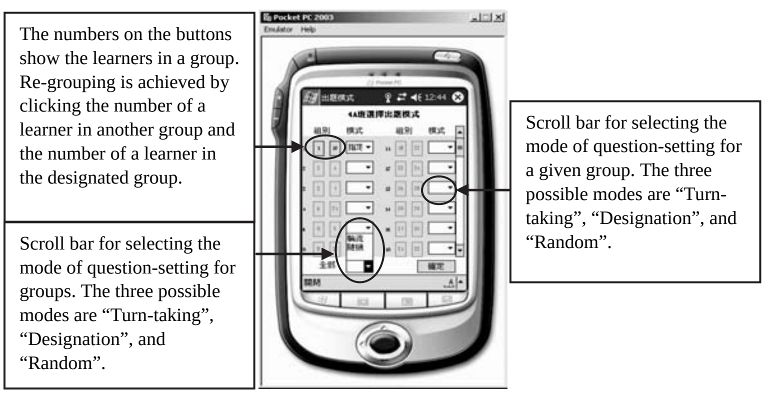
Once the teacher believes that all of the groups have reached learning status 3 (in which learners have a good understanding of the concept of fraction equivalence) then the mode of question-setting can be changed to "Random" for the entire class, which means that the role of question-setter is assigned randomly by the computer system. The "Random" mode provides learners with the opportunity to explore the concept further in a relaxed mode of inquiry, which helps to consolidate the learning outcomes.

Generating Procedural Knowledge of Adding Fractions with Unlike Denominators

The ultimate goal of the collaborative mobile learning environment is to support learners to generate procedural knowledge of adding fractions with unlike denominators. With a solid knowledge of fraction equivalence, learners will enter a proximal zone in which they can develop the procedural knowledge of adding fractions with unlike denominators.

The aforementioned three spaces of the model of affordances equip learners with a comprehensive understanding of fraction equivalence. The mobile version of the CT serves as a platform to stimulate the cognitive elaboration of learners to help them to derive the algorithm of adding fractions with unlike denominators. By putting learners through

Figure 9. Teacher's interface for the re-grouping of learners and the changing of the question-setting mode



the steps of expanding the designated fractions to reach a common denominator and then adding the two expanded fractions with the support of graphical representation, the CT fosters learners to associate learning how to add fractions with the known concept of fraction equivalence.

CONCLUSION

The desktop version of a Web-based cognitive tool (CT) that supports the acquisition of procedural knowledge of mathematical fractions is migrated to a mobile version to increase its learning effectiveness by taking advantage of collaborative learning and mobile learning. Using the design-based research approach, a theory-driven design of a collaborative mobile learning environment is established based on two empirical evaluation research studies.

The mobile version CT is designed as a model of affordances for learning about fraction equivalence with three spaces that offer learners profitable learning opportunities. The first space comprises a learner-controlled animation of partitioning to help learners to develop the part-whole concept and understand the inverse relationship between the number of parts and the size of a part of a unit. The second space comprises an animation that compares the equivalence of fractions and features a random display of fraction bars in non-comparable representations to help learners to represent fractions in comparable forms to determine their equivalence. The third space comprises the hypothesis-testing interface $\frac{a}{b} = \frac{a \times c}{b \times d}$ for finding equivalent fractions, which helps learners to develop the capability to compute equivalent fractions by adjusting parameters *c* and *d*.

The designed learning environment offers profitable opportunities for learners to share knowledge with their peers, provides pedagogical designs for teachers to mediate cognitive elaboration, and allows teachers to organize learners into heterogeneous groups. Two pedagogical designs are suggested to realize these pedagogical benefits. The first aims to promote situations in which reflection and cognitive elaboration in the collaborative mobile learning environment take place, and involves grouping learners into pairs to engage in in-depth discussion about the learning activities. Learners are encouraged to learn together both as peers through tasks that are designed for collaborative engagement, and as partners through friendly competition to trigger in-depth cognitive elaboration.

Reciprocal tutoring is a key strategy for helping learners resolve cognitive conflict, and thus the second pedagogical design aims to promote an environment for reciprocal tutoring in a mobile technology supported classroom. Teachers in the designed learning environment act as facilitators to mediate and promote the sharing of knowledge among the class, and can reorganize learning pairs using the mobile device to promote reciprocal tutoring.

The architectural design of the artifacts and the pedagogical design of the learning activities enable learners to develop a concept of fraction equivalence using a collaborative interactive approach. With a solid knowledge of fraction equivalence, learners will enter a proximal zone in which they can develop and generate procedural knowledge of adding fractions with unlike denominators.

The use of mobile technology to encourage collaborative learning is a promising research direction that deserves further study, especially in terms of its effect on classroom learning environments. We have begun a case study of learning the concept of fraction equivalence in this collaborative learning environment, and further large-scale studies to investigate whether learners recognize and reappraise anomalies and the way in which they attempt to resolve cognitive conflict in the learning process will be attempted after the completion of the pilot case study.

REFERENCES

Akhras, F. N., & Self, J. A. (2000). Modeling the process, not the product, of learning. In S. P. Lajoie (Ed), *Computers as cognitive tools, volume two: no more walls* (pp. 3-28). Mahwah, NJ: Lawrence Erlbaum Associates.

Anderson, J. R. (1976). *Language, memory, and thought*. Hillsdale, NJ: Erlbaum.

Bell, P. (2004). On the theoretical breadth of design-based research in education. *Educational Psychologist*, *39*(4), 243-253.

Cantor, G. N. (1983). Conflict, learning, and Piaget: comments on Zimmerman and Bloom's "Toward an empirical test of the role of cognitive conflict in learning." *Developmental Review*, *3*, 39-53.

Chan, T. W., & Chou, C. Y. (1997). Exploring the design of computer supports for reciprocal tutoring. *International Journal of Artificial Intelligence in Education, 8*, 1-29.

Clark, A. (1997). *Being there: Putting brain and world together again*. Cambridge, MA: MIT press.

Derry, S. J., & Lajoie, S. P. (1993). A middle camp for (un)intelligent instructional computing: an introduction. In S. P. Lajoie & S. J. Derry (Eds), *Computers as cognitive tools* (pp. 1-11). Hillsdale, NJ: Lawrence Erlbaum Associates.

Design-Based Research Collective. (2003). Design-based research: An emerging paradigm for educational inquiry. *Educational Researcher, 32*(1), 5-8.

Dillenbourg, P. (1999). Introduction: What do you mean by "collaborative learning?" In P. Dillenbourg (Ed), *Collaborative learning: cognitive and computational approaches* (pp. 1-19). Amsterdam: Pergamon.

Doherty, R. W., Hilberg, R. S., Pinal, A., & Tharp, R. G. (2003). Five standards and student achievement. *NABE Journal of Research and Practice*, *1*(1), 1-24.

Fischer, F., Bruhn, J., Gräsel, C., & Mandl, H. (2002). Fostering collaborative knowledge construction with visualization tools. *Learning and Instruction, 12*, 213-232.

Gibson, J. J. (1979). *The ecological approach to visual perception*. Boston: Houghton Mifflin.

Hoadley, C. M. (2004). Methodological alignment in design-based research. *Educational Psychologist*, *39*(4), 203-212.

Huinker, D. (1998). Letting fraction algorithms emerge through problem solving. In L. J. Morrow & M. J. Kenney (Eds), *The teaching and learning of algorithms in school mathematics* (pp. 170-182). Reston, VA: NCTM.

Janney, R. W. (1999). Computers and psychosis. In J. P. Marsh, B. Gorayska & J. L. Mey (Eds.), *Humane interfaces: questions of method and practice in cognitive technology* (pp. 71-79). Amsterdam: Elsevier Science.

Johnson, D. W., & Johnson, R. T. (1999). *Learning together and alone: cooperative, competitive and individualistic learning (5th Ed.)*. Boston: Allyn & Bacon.

Kommers, P. A. M., Jonassen, D. H., & Mayes, J. T. (1992). *Cognitive tools for learning*. Heidelberg: Springer-Verlag.

Kong, S. C. (2001). Interactive Web-based learning tools for knowledge construction on subject-specific contents. In K. S. Volk, W. M. So & P. T. Gregory (Eds.), *Science & Technology Education Conference 2000 Proceedings* (pp. 374-386). Hong Kong: Science & Technology Education Conference.

Kong, S. C., & Kwok, L. F. (2002). Modeling a cognitive tool for teaching the addition/subtraction of common fractions. *International Journal of Cognition and Technology*, *1*(2), 327-352.

Kong, S. C., & Kwok, L. F. (2003). A graphical partitioning model for learning common fractions: designing affordances on a web-supported learning environment. *Computers and Education*, *40*(2), 137-155.

Kong, S. C., & Kwok, L. F. (2005). A cognitive tool for teaching the addition/subtraction of common fractions: a model of affordances. *Computers and Education*, *45*(2), 245-265.

Lee, G. H., Kwon, J. S., Park, S. S., Kim, J. W., Kwon, H. G., & Park, H. K. (2003). Development of an instrument for measuring cognitive conflict in secondary-level science classes. *Journal of Research in Science Teaching, 40*(6), 585-603.

Limón, M. (2001). On the cognitive conflict as an instructional strategy for conceptual change: a critical appraisal. *Learning and Instruction*, *11*, 357-380.

Niemi, D. (1996). Assessing conceptual understanding in mathematics: Representations, problem solutions, justifications, and explanations. *Journal of Educational Research*, *89*(6), 351-363.

Pea, R. D. (1985). Beyond amplification: using the computer to reorganize mental functioning. *Educational Psychologist*, *20*(4), 167-182.

Pitkethly, A., & Hunting, R. (1996). A review of recent research in the area of initial fraction concepts. *Educational Studies in Mathematics*, *30*, 5-38.

Roschelle, J. (2003). Keynote paper: Unlocking the learning value of wireless mobile devices. *Journal of Computer Assisted Learning*, *19*, 260-272.

Roschelle, J., & Teasley, S. D. (1995). The construction of shared knowledge in collaborative problem solving. In C. O'Malley (Ed.), *Computer-supported collaborative learning* (pp. 145-168). Berlin: Springer-Verlag.

Sharples, M., Taylor, J., & Vavoula, G. (2005). Towards a theory of mobile learning. *Proceedings of the mLearn 2005 Conference* (pp. 9). Cape Town: mLearn 2005.

Slavin, R.E. (1996). Research on cooperative learning and achievement: What we know, what we need to know. *Contemporary Educational Psychology*, *21*(1), 43-69.

Tall, D. (1977). Conflicts and catastrophes in the learning of mathematics. *Mathematical Education for Teaching*, *2*(4), 2-18.

Wittrock, M. C. (1979). The cognitive movement in instruction. *Educational Researcher, 8*(2), 5-11.

Wittrock, M. C. (1986). Students' thought processes. In M. C. Wittrock (Ed.), *Handbook of research on teaching* (pp. 297-314). NY: Macmillan Publishing Co.

Wong, W. K., Chan, T. W., Chou, C. Y., Heh, J. S., & Tung, S. H. (2003). Reciprocal tutoring using cognitive tools. *Journal of Computer Assisted Learning*, *19*, 416-428.

KEY TERMS

Affordances: Making available profitable spaces in which learners can interact in ways that meet their needs.

Cognitive Conflict: A perceptual state in which one notices the discrepancy between an anomalous situation and a preconception.

Cognitive Elaboration: The process of forming associations between new information and prior knowledge.

Collaborative Learning: A process that encourages learners to participate in coordinated and synchronous learning activities with a number of other learners.

Cognitive Tools: Mental and computational devices that can support, guide, and mediate the cognitive processes of learners.

Mobile Learning: The use of mobile technology for learning and teaching.

Procedural Knowledge: Knowledge that guides the performance of certain tasks in the absence of the knowledge that underlies the performance of the procedure.

Chapter XVIII
Design of an Adaptive Mobile Learning Management System

Hyungsung Park
Korea National University of Education, Korea

Young Kyun Baek
Korea National University of Education, Korea

David Gibson
The University of Vermont, USA

ABSTRACT

This chapter introduces the application of an artificial intelligence technique to a mobile educational device in order to provide a learning management system platform that is adaptive to students' learning styles. The key concepts of the adaptive mobile learning management system (AM-LMS) platform are outlined and explained. The AM-LMS provides an adaptive environment that continually sets a mobile device's use of remote learning resources to the needs and requirements of individual learners. The platform identifies a user's learning style based on an analysis tool provided by Felder & Soloman (2005) and updates the profile as the learner engages with e-learning content. A novel computational mechanism continuously provides interfaces specific to the user's learning style and supports unique user interactions. The platform's interfaces include strategies for learning activities, contents, menus, and supporting functions for learning through a mobile device.

INTRODUCTION

The rapid advancement of the global information infrastructure, mobile informational technologies, and intelligent applications is leading to a change of educational paradigm. The new paradigm is evolving similar to the way that "distance learning" evolved into "e-learning." Now e-learning is changing into m-learning (mobile learning) and is providing new possibilities for education.

Among those possibilities are increased ability to promote student motivation through personalization and a change from teacher-centered teaching to learner-centered learning. Adaptive m-learning can support these possibilities by considering and using a learner's diverse variables, such as abilities, attitudes, and learning styles, to promote effective learning and place the learner at the center of a more personalized experience.

Every teacher has witnessed how some students prefer visual information while others are surprised and perplexed when complex diagrams are given. Although one student may be weak in a speed test, he or she might understand a discourse more deeply than another student and be able to submit substantial and excellent reports. Learners also vary in their backgrounds and experience, and possess a diversity of abilities that cause them to learn in different ways. They are unique in their personalities and values, for example. In addition, they develop individual preferences for learning environments that support their favored modalities of learning. In general, students exhibit a wide variety of unique blends of strengths and weaknesses resulting in classrooms with a wide diversity of talents that need to be developed.

The benefits of personalizing learning are well documented in the literature on differentiation of learning (Brimijoin 2003; Stevens 1999; Tomlinson & National Association for Gifted Children 2004) and are also easy to illustrate with an example. If the classroom has as few as two different kinds of learners and only one kind of instruction used, there will be a "best and worst fit" among the students. If the same instruction is used repeatedly, then one of the students will be systematically denied access to the most effective instruction.

Understanding the different ways that children learn, interact with, and process information can help teachers modify instruction so that all students have an equal opportunity to succeed (Theroux, 2004). In order for teaching to be an intentional and planned activity that supports each student's academic success, it is necessary to accept and utilize each learner's features to foster the most effective learning. It follows that teachers, learning devices, and instructional programs that provide a variety of learning approaches have a greater chance of offering appropriate challenges to every student in the learning environment. However, with highly portable m-learning, the teacher's role needs to shift to the device.

When the learning environment involves mobile devices, the variety of learners' background, abilities, and learning styles are expected to be more diverse than in a traditional classroom environment. This is true because the mobile device can be picked up and used by anyone at anytime, with or without a teacher present. The handheld learning environment thus needs a great deal of adaptability. It must be able to support independent learning without expecting a teacher's support and guidance. As we envision it, the mobile device itself can play an adaptive role that shapes the learning environment on the basis of a learner's preferred style.

To capture this idea, Park developed a prototype adaptive mobile learning management system (AM-LMS) which assesses a user's learning style, creates a learner profile, and then provides content based on decisions the learner makes while interacting with the content, continuously updating the learner profile. The chapter presents background, rationale and a functional overview of the AM-LMS.

MOBILE LEARNING

Mobile learning is based on wireless Internet connections and uses devices such as notebook computers, cellular phones, personal communication system (PCS) phones, and personal digital assistants (PDAs). The important features of mobile devices are their portability, immediacy, individuality, and accessibility; features which are bringing about a change of paradigm in approaches to teaching (Shostsberger & Vetter, 2000).

Dye et al. (2003) define m-learning as "learning that can take place anytime, anywhere with the help of a mobile computer device. The device must be capable of presenting learning content and providing wireless two-way communication between teacher(s) and student(s)" (p. i). Figueiredo and Chabra (2002) emphasize device flexibility; not only does m-learning offer the ability to receive learning anytime, anywhere, but as important, on any device. Harris (2001) adds to the definition by defining m-learning as that point at which mobile computing and e-Learning intersect. The term "mobile learning" in this chapter will be defined broadly as a form of learning delivered through mobile devices such as mobile phones, PDAs, smart phones, tablet PCs and similar devices combined with e-learning content.

Mobile Learning Environment

Mobile devices are a familiar part of the lives of most teachers and students. They offer the opportunity to embed learning in a natural environment (Schwabe & Goth, 2004) and enable learning that is independent of time and location constraints and with increasingly customized contents (Abfalter et al., 2004). Currently there are increasing efforts to apply mobile technology to learning (see, for example, Gay, Reiger, & Bennington, 2001; Hoppe, Joiner, Milrad et al., 2003; Kristiansen, 2001; Lundin, Nulden, & Persson, 2001; Schwabe & Goth, 2005; Sharples, Corlett, & Westmancott, 2002). Sharples et al. (2002) point out that there may be a particular opportunity for mobile learning outside the traditional formal learning settings.

The last decade has seen far-reaching changes in living, learning, working, and collaboration, fundamentally influenced by information and communication technologies, specifically the World Wide Web. Projecting 10 years into the future we may ask what the new impact of wireless and mobile technologies will be. We should take up the challenge that the future is not "out there" to be discovered, but has to be invented and designed to meet new needs and possibilities that emerge as wireless and mobile technologies become widely available (Fisher & Konomi, 2005).

A mobile learning environment requires a wireless Internet service that provides content upon a request by a client who holds a mobile device (Figure 1). To supply learning content to the wireless Internet, Web services are provided through a WAP Gateway, which enables regular communication with mobile devices. The WAP Gateway allows both WAP and TCP/IP protocol. Through the network-enabled WAP Gateway, a mobile learning environment can provide students and teachers with the opportunity to obtain any and all class-related material on a handheld computer, such as a Palm Treo, through a simple process of point-and-connect using infrared.

Landers (2005), the forum administrator for 'From e-learning to m-learning,' presented options created by mobile learning methods in Europe as

Figure 1. Systematic structure of wireless Internet

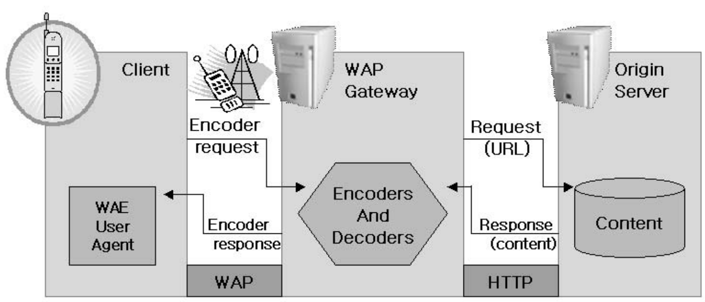
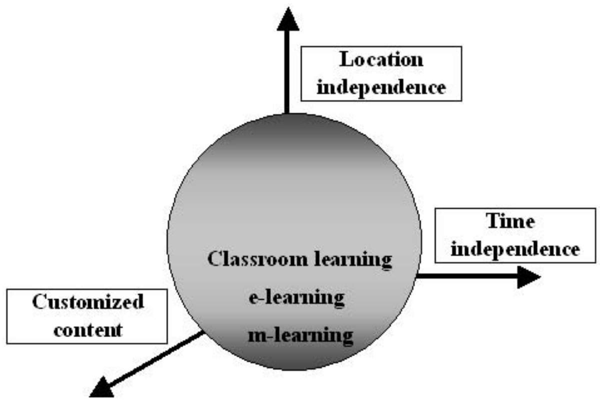


Figure 2. Expansion of mobile learning environment (Abfalter et al., 2004)



follows: first, learning content can be separated from mobile media (Ketamo, 2005). Second, content can be used to help students study a regular curriculum in the classroom (Vahey & Crowford, 2005). Third, they can be integrated into web-based learning and web based instruction (Heoyoung, 2002). Fourth, learning communities can share material and undertake self-directed study. Fifth, mobile technology can be used by students when taking trips or participating in field activities (Gay et al., 2002). Sixth, the mobile tools can be the main materials while others support the mobile materials.

Mobile learning with an adaptive learning management system supports learning that is independent of both time and location through contents provided on demand to the wireless device. The intersection of mobile computing and e-learning also provides strong search capabilities; rich interactions with users; powerful support for effective learning; and performance-based assessment. Wireless communication expands the learning space beyond the traditional classroom to unlimited cyber space. Abfalter et al. (2004) for example, explains that in a mobile learning environment, "Teaching and learning are no longer confined to time and place. With mobile communication technologies, the time and physical boundaries of the traditional classroom are being expanded (p. 2)." Figure 2 shows the independent characteristics of mobile learning.

Research Trends in Mobile Learning

Three primary trends can be found in the research on mobile learning: research on the interface contained in learning contents, research on the learning management system, and research on mobile learning patterns based on ubiquitous computing.

Research on the interface considers the special qualities of mobile devices: small screen size, slow text input facilities, small storage capacity, limited battery life, low bandwidth network capabilities and slow CPU speed. In particular, the limitation of small screen directly affects the user's learning process and behavior. Recent studies (Buyukkoten et al., 2000; Kawachiya & Ishikawa, 1999) on the effect of screen size on completing browser related tasks for example, show that mobile users tend to follow links less frequently than traditional Internet users (Antonellis et al., 2005).

Another theme of the research concerns learning management and supporting systems. Studies have found that teachers use mobile devices for attendance reporting, reviewing student marks, general access of central school data, and managing their schedules more effectively (Laura, 2006). Also, handheld computers have been found to bring important benefits to schools by assisting administration, supporting classroom management, and enabling personal and group learning (Perry, 2003). Research studies by the

Becta PDA (personal digital assistant) project (2003) were focused on two aspects: managing teacher's workload and supporting teaching and learning on PDA use. They found that PDAs have considerable potential for improving a teacher's management and presentation of information. At the same time, PDAs were found to bring important benefits to schools by enabling individualized learning and group learning.

A third type of m-learning research focuses on ubiquitous computing, in which a huge number of tiny computers are embedded into an invisible part of the fabric of everyday life—in watches, microwave ovens, cars, and clothes. Recently m-learning researchers have been proposing adapted learning contents for a variety of learning styles, joined with the idea of ubiquitous computing (Sakamura & Koshizukea, 2005). For example, Shindo et al. (2003) introduces the idea of a digital ubiquitous museum that embeds ubiquitous learning materials and tools into daily living environments. Also, Deng et al. (2005) proposed using wireless and mobile devices to support academic conferences with an "ask the author" application. Schwabe & Goth (2005) reports on the design of a mobile orientation game in a university setting. Gay et al. (2001) studied the use of mobile computing to enhance field study. MOBIlearn, a major European research project, is focusing on the context-aware delivery of content and services to learners with mobile devices (Lonsdale, Nudin & Persson, 2004). With these kinds of research efforts, it seems timely to suggest an adaptive "learning style" architecture for mobile learning device interfaces to a learning management system.

LEARNING STYLES

Felder (1996) and many others have made the point that students have different learning styles—characteristic strengths and preferences in the ways they take in and process information. Some students tend to focus on facts, data, and algorithms; others are more comfortable with theories and mathematical models. Some respond strongly to visual forms of information, like pictures, diagrams, and schematics; others get more from verbal forms-written and spoken explanations. Some prefer to learn actively and interactively; others function more introspectively and individually. Keefe (1982) explained that learning styles are the cognitive, affective, and psychological traits that serve as relatively stable indicators of how learners perceive, interact with, and respond to the learning environment. Dunn (2000) explained that learning style is the way a person processes, internalizes, and studies new and challenging material. The cornerstone of the learning style theory is that most people can learn, and each individual has his or her own unique ways of mastering new and difficult subject matter.

Kolb's (1984) point of view is that teaching with learning styles is a way shaping and intensifying learning by making the environmental demands of learning tasks coincide with individual preferences. Felder and Silverman (1996) defined learning styles as the process of acquiring and controlling information with the traits and preferred ways in which students study. Della-Dora and Blanchard (1979) also believe that "learning style can be defined as a personally preferred way of dealing with information and experience that crosses content areas." Many researchers have grasped the importance of the preference factor to improve learning (Kruzich, Friesen & Van Soest, 1986).

In brief, learning style can be described as the general characteristics of an individual's intrinsic procedures of information processing that lead to unique behavioral patterns, which tend to be durable and stabile in a variety of learning environments.

Index of Learning Styles

The Index of Learning Styles is an online instrument used to assess learning preferences on four dimensions (active/reflective, sensing/intuitive, visual/verbal, and sequential/global) of a learning style model formulated by Richard M. Felder and Linda K. Silverman. Richard M. Felder and Barbara A. Soloman of North Carolina State University developed the instrument.

The Felder and Soloman (2005) Index of Learning Style (ILS) provides critical guidance for determining the type of content when individual students are learning with the adaptive mobile learning management system (AM-LMS). The ILS analysis contributes to the AM-LMS system in two aspects. First, the analysis suggests one of sixteen learning styles identified by the Felder and Soloman model and second, those styles are sufficient to provide relevant learning contents according to the various learners' needs in the process of learning. On this theoretical foundation, the AM-LMS platform is structured to provide an interface for adaptive contents.

Four Dimensions of Learning Style

In a report by Felder and Silverman (1988) a model of learning style was originally described and is now classified by four dimensions (2005). The first dimension is related to the question, what type of information does the student preferentially perceive: sensory-sights, sounds, sensations, or intuitive-memories, ideas, insight? The second dimension concerns effective perception and is the answer to the question, through which modality is sensory information most effectively perceived: *visual*—pictures, diagrams, graphs, demonstrations, or *verbal*—written and spoken words and formulas? The third dimension is the pattern of processing information and the answer to the question, how does the student prefer to process information: *actively*—through engagement in a physical activity or discussion, or *reflectively*—through introspection? The fourth dimension is the type of the progress to understanding: How does the student progress toward understanding: *sequentially*—in a logical progression of small incremental steps, or *globally*—in large jumps, holistically? (Felder, 1995).

It is convenient to set up pairs of opposites on the ends of continua representing the complex mental processes by which perceived information is converted into knowledge. In what follows, four such pairs of opposites further refine the four learning style dimensions.

Active and Reflective Learners

An "active" learner is a person with a natural tendency toward active experimentation more than toward reflective observation. This type of learner tends to solve problems through discussion and group work and is extroverted. A "reflective" learner is conversely, a person with a natural tendency toward reflective observation more than toward active experimentation. This type of learner tends to solve problems through self-evaluation and reflection and is introverted.

Sensing and Intuitive Learners

"Sensing" learners prefer to accept information in a well-structured order. They tend to be concrete and methodical and they prefer facts, data, and experimentation. They are patient with detail but dislike complications and they rely more on memorization as a learning strategy. They are more comfortable learning and following rules and standard procedures than using their intuition. "Intuitive" learners, on the other hand, prefer to organize information in their memory according to their own rules. They tend to be abstract as well as imaginative and deal better with principles, concepts, and theories. They are apt to be bored by details and they welcome complications.

Visual and Verbal Learners

"Visual" learners prefer that information be presented in pictures, diagrams, flow charts, time lines, films, and demonstrations—rather than in spoken or written words. Meanwhile, "verbal" learners prefer spoken or written explanations to visual presentations.

Sequential and Global Learners

"Sequential" learners absorb information and acquire understanding of material in small, connected chunks, whereas "global" learners take in information in seemingly unconnected fragments and achieve understanding in large holistic leaps.

Index of Learning Style Questionnaire

The Index of Learning Style (ILS) is a 44-item, forced-choice instrument first developed in 1991 by Richard Felder and Barbara Soloman to assess learning preferences on the four scales of the Felder-Silverman model. Currently, the Felder and Silverman theory (2005) categorizes an individual's preferred learning style by a sliding scale of four dimensions: sensing-intuitive, visual-verbal, active-reflective and sequential-global (Table 1).

The ILS questionnaire consists of 44 questions and each with two possible answers: a or b. All questions are classified corresponding to the four pairs in the Felder and Silverman learning style theory. Each dimension has 11 questions and the 16 learning styles are classified based on the scores earned on each dimension. An example of a questionnaire based on the Felder and Silverman index of learning style is presented in a screen shot of the AM-LMS in Figure 3.

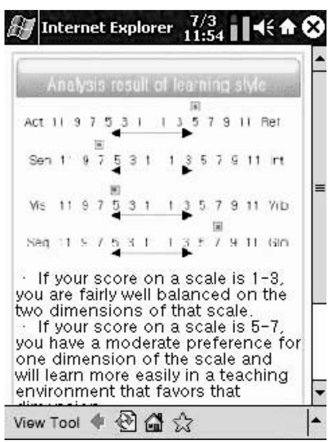
The scores earned on the dimensional scales of the questionnaire are explained as follows:

- If the score on any scale is 1-3, it means that the student is fairly well balanced on the two dimensions of that scale.
- If the score on a scale is 5-7, it means a moderate preference for one dimension of the scale and the students will learn more easily in a teaching environment that favors that dimension.
- If the score on a scale is 9-11, it means a very strong preference for one dimension of the scale. The student may have difficulties in learning environment that does not support that preference.

Table 1. Felder's learning dimensions (Carver et al., 1999)

Definitions	Dimension		Definitions
Do it	Active	Reflective	Think about it
Learn facts	Sensing	Intuitive	Learning concepts
Require Pictures	Visual	Verbal	Require reading or lecture
Step by Step	Sequential	Global	Big picture

Figure 3. Learning style results of ILS



The Classification of Learning Style for Adaptation

The AM-LMS system classifies the learning style of the user by using the ILS assessment to build a profile of the learner. It then constructs adaptive contents for the learning process. It initially discriminates the user's learning style by a combination of the items selected with weak preference and the items with strong preference. Those learning styles are then used as criteria to provide adaptive contents for the learning process, the application selects from among the 16 possible adaptive content types that results from 4 pairs of opposites (8 dimensions of the learning style typology), as shown in Figure 4.

Here are the rules used by the AM-LMS for constructing adaptive contents based on learning styles:

- **Active:** An active learner asks questions frequently and is given answers accordingly. In addition, the system provides these learners with a discussion room, bulletin board, and chat room.
- **Reflective:** A reflective learner needs to contemplate the topic of a subject in advance before starting a lesson. During the lessons, the learner summarizes his or her learning and reviews the whole learning process. The AM-LMS system provides this style of learner with relevant tools, and confirms prerequisite knowledge before the lesson, then provides reflective learners with links for related learning, materials for downloading and performing evaluations.
- **Sensing:** A sensing learner needs an example of the facts, demonstrations and practical materials. The AM-LMS provides easily located learning contents and summarizes and structures the learning contents for sensing learners.
- **Intuitive:** An intuitive learner needs to be provided with concepts, abstract ideas, demonstrations, and theories before concrete examples. The AM-LMS provides a sequential order of exposition followed by examples to best fit the intuitive learner.
- **Visual:** A visual learner needs, and the AM-LMS provides, pictures, graphs, diagrams, flow charts, schematics, demonstrations, concept maps, color notes, and slides with multimedia.
- **Verbal:** For verbal learners, the AM-LMS presents content primarily as text and audio.
- **Sequential:** For sequential learners, the AM-LMS structures material in a logical, step by step, orderly outline.
- **Global:** A global learner needs to see the big picture before the details in order to view more of the context of a subject. The AM-LMS presents the big picture of the course and all the links are made available

Figure 4. Combinations of eight learning style dimensions to create 16 learning styles

[1] Active Sensing Visual Sequential	[5] Active Intuitive Visual Sequential	[9] Reflective Sensing Visual Sequential	[13] Reflective Intuitive Visual Sequential
[2] Active Sensing Visual Global	[6] Active Intuitive Visual Global	[10] Reflective Sensing Visual Global	[14] Reflective Intuitive Visual Global
[3] Active Sensing Verbal Sequential	[7] Active Intuitive Verbal Sequential	[11] Reflective Sensing Verbal Sequential	[15] Reflective Intuitive Verbal Sequential
[4] Active Sensing Verbal Global	[8] Active Intuitive Verbal Global	[12] Reflective Sensing Verbal Global	[16] Reflective Intuitive Verbal Global

as flexible options to ensure free movement by the global learner.

Figure 5 shows a screen of content presented adaptively according to a corresponding learning style. The display in the left of Figure 5 is constructed for the ASVG (Active, Sensing, Verbal, Global) style among 16 styles presented in Figure 4. The display in the right of Figure 5 is constructed for the ASVS (Active, Sensing, Visual, Sequential) style.

User Interface for Mobile Learning

The user interface in mobile learning is different from other computing interfaces in its size of display. Display sizes range from 100 x 80 pixels of mobile phones to 240 x 320 pixels for personal digital assistants. That small display area, its color, and the amount of displayable information are major points which need to be considered when adapting contents.

The AM-LMS uses several strategies to deal with the smaller display. First, the amount of information has been reduced so that the mobile device can accept it even when the learner is traveling. The wireless Internet also limits the amount of information that can be sent and received. Thus, the scroll has been minimized in AM-LMS content. Second, the user interface avoids asking the learners for a lot of input because of the limitation of input devices on mobile devices. Thus simple response patterns such as pointing and clicking without moving around text areas has been adopted. Third, vertical navigation within a same topic has been confined to one or two hyperlink depths. Fourth, sentences are simple as opposed to compound and complex. Fifth, the location and the topic in current learning are displayed so that the learner can understand what he or she is doing under any topic.

To deliver the interface adapted to each learner's learning style, the AM-LMS analyzes past performance, including the initial survey information and the learner's subsequent choices and selections, and then generates new content and presentations based on that analysis. For example, reflective learners need to review their learning process. When given reflective opportunities, they can manage their own learning based on their judgment of what is working for them. Reflective users navigate their learning through continuous adjustments in their learning behaviors based on questions such as "What are the key characteristics of this material that help me understand?" "What is the criteria for mastery of what I am doing?" and "How am I conceptualizing this problem?" Schön (1983). Thus, in the interface for the reflective learners, elements prompting reflective thinking are introduced in display forms, inducing and stimulating questions which help them understand what they are doing and the relationships among previous and current topics.

Figure 5. Contents adaptive to learning styles (ASVG, ASVS)

The AM-LMS establishes similar strategies of user interface for each of the eight elements of learning style (Table 2).

An outstanding question for future research concerns the peculiarities of mobile device interaction that may affect the applicability of the four dimensional model. For example, a highly visual learner may be more impacted by the screen limitations of the mobile device than a textual learner. Research designs are needed that explore the learning effectiveness of handhelds with and without adaptive LMS as well as the degree of change in a student's choices of learning styles over time. Comparisons of mobile adaptive versus traditional learning management systems and traditional systems with learning styles adaptability are needed.

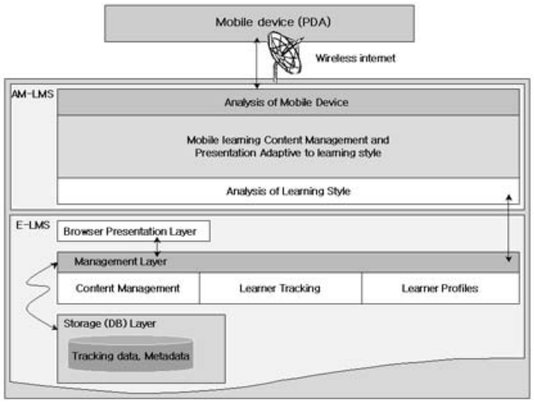
STRUCTURE OF THE ADAPTIVE MOBILE LEARNING MANAGEMENT SYSTEM

As the previous sections have hopefully made clear, the adaptive mobile learning management system (AM-LMS) is a platform for providing learning contents adapted to a learner's learning style in a mobile learning environment. The AM-LMS manages the whole learning process, monitors the progress of a learner and presents learning contents adaptive to the user's learning style, including the potential change in learning preferences that may appear over time. The system plays its role following each learning stage by analyzing learning progress and giving feedback. Figure 6 shows the functional structure of the AM-LMS.

Table 2. Design strategies according to learning style (based on Felder & Silverman, 1988)

Learning Style	Characteristic	AM-LMS Interface Strategy
Active	Has outward character and prefers problem solving through discussion and cooperative/group work	Provides bulletin board and chatting room
Reflective	Has inward character and prefers problem solving through self assessment and reflections	Provides links to other references, downloadable materials, and performance evaluation materials
Sensing	Prefers the ways for understanding in order in organized pattern	Provides well organized and summarized information in balanced location in the display
Intuitive	Prefers the ways for organizing information with his or her freedom for easy memorizing	Describes text information in narration format
Visual	Prefers visual information such as pictures, graphs, drawings	Provides pictures and graphs with explanation
Verbal	Prefers audible information	Presents text with audible information
Sequential	Prefers sequential and structural ways of learning	Organizes content in order and makes navigation exploratory
Global	Prefers the ways of understanding content as a whole and overall perspective with freedom	Makes navigation free so that content can be selected with learner's will

Figure 6. Structure of the adaptive mobile-learning management system



The AM-LMS consists of (1) Access device analysis, (2) Learning style analysis, (3) Mobile content management and presentation adaptation (Figure 6). The AM-LMS works on top of an e-learning platform that is responsible for the general management of learners and learning contents. The extra mobile learning management system analyzes mobile devices, converts e-learning contents into the m-learning contents suitable for each learner and presents the contents adaptively on screens and other devices. A learner interacts through question items and other products, which are provided to the WAP Gateway through a mobile device that transmits them to the AM-LMS.

The AM-LMS system aims to provide learners with specific contents suitable to their learning styles so that they can learn in ways that are based on their own needs. Learners at a distance cannot expect to get a teacher's support and guidance as they can in classroom settings, thus the AM-LMS system takes many possible characteristics into consideration including learners' backgrounds, capabilities, and learning styles. The AM-LMS has a stand-alone capability for building a rich and responsive learning environment without any teacher's support.

Analyzing Learning Style

The learning style analysis module evaluates the learner and stores the evaluation into a personal profile. When the learner comes back later, the profile provides basic data so that the AM-LMS can supply adaptive contents on the basis of the variously designed learning styles materials in a database. Figure 7 shows a screen to analyze the learning style of the user through a PDA.

Analysis Module

When the mobile device accesses the system, the analysis module automatically determines the type of device in use and transfers that information to the mobile learning management and presentation adaptation modules. This analysis module determines the device type, running

Figure 7. Screens analysis of learning style

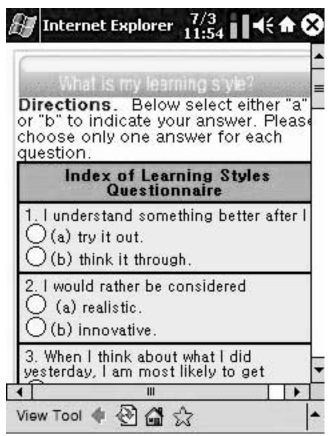
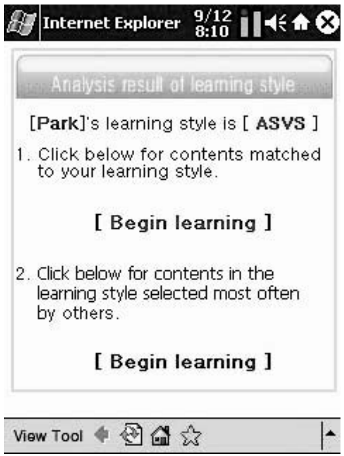


Table 3. Sample code to identify a browser

```
If (InStr(Request.ServerVariables(“HTTP_USER_AGENT”), “Windows CE”)) Then
“ REDIRECT TO CODE FOR HANDHELD
Else
“ YOUR NORMAL WEB SITE
End If
```

Figure 8. Choice of learning process through the analysis results of learning styles



environment, features, and operating system of a mobile device that is accessing learning content by using the information in the header. In order to identify a handheld device, a Windows CE browser also identifies itself via the HTTP request header (Table 3).

Managing and Adapting Mobile Contents

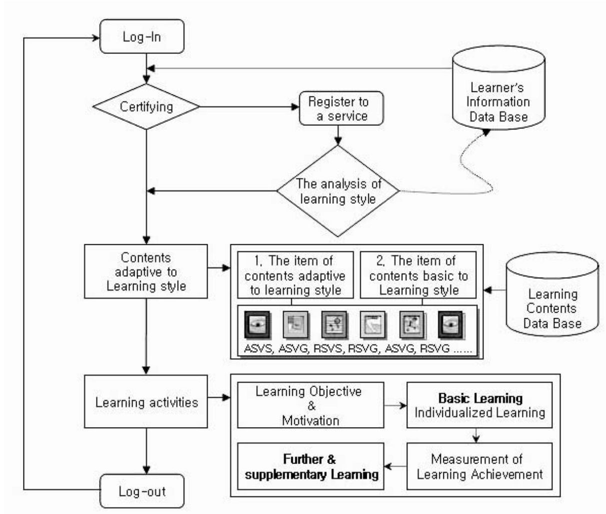
A PDA’s screen size is limited to 240 by 320 pixels and a cellular phone and smart phone have even smaller screen sizes. In addition to smaller amounts and types of information, the AM-LMS module presents two elements in response to the learning progress selected by a learner:

1. Learning path matched to the learner’s learning style. This link functions as an opening for packaging and delivering contents suitable for each learner’s learning style. Figure 8, the first screen for the learner, displays the learner’s learning style on the first row and has options to get into either the learning path specific to his/her learning style or a universal learning path which most learners has taken before this session.
2. The learning path taken by most learners. This link functions as an opening for packaging and delivering contents preferred or taken by most learners. This link also is a path to frequently presented contents.

Flow of Adaptive Mobile Learning

In the AM-LMS, a learner advances his or her own learning by making frequent decisions. Then the learner is provided with learning content according to the analyzed learning style. Learning in the AM-LMS flows as presented Figure 9.

Figure 9. Flow of learning in AM-LMS



The learner enters the learning environment through the process of certification on the login screen (Figure 9). If the learner has his or her own learning style analyzed previously, he or she enters the learning process that matches his or her learning style. Otherwise, he or she advances into the learning process after the learning styles analysis. The AM-LMS system periodically keeps track of the progress of learners' activities and continuously stores the learners' profile and tracking information.

CONCLUSION

The AM-LMS design introduces the theoretical background for the development of an adaptive mobile learning management system that presents content in response to the learning style of a user. It enlarges the concept of adaptive learning system approaches in traditional LMS systems to include mobile devices. The Felder and Silverman (2005) learning style theory provides the platform with a framework to analyze a user's learning style and present contents adapted to that style. The system can thus promote individualized learning and learner-centered education while taking advantage of the unique features of mobile devices and network-based learning environments.

The AM-LMS is an ongoing research project. Learner variables such as cognitive style, task commitment and others will be added to the system in the future so that it can be more adaptive. At this point, interface research on content size, navigation, and display techniques are urgently needed in the area of mobile devices. In addition, issues such as cognitive overloading for the learners on mobile devices have not yet been adequately studied, although there are several related research findings for the Web and other information technologies. The AM-LMS conceptual model is thus expected to adopt more variables resulting from future research studies and will hopefully contribute to strong and effective ways of achieving individualized learning on mobile devices.

REFERENCES

Abfalter D., Mirski P. J., & Hitz, M. (2004). Mobile learning—Knowledge enhancement and vocational training on the move. In *The 5th European Conference on Organizational Knowledge, Learning and Capabilities.* Innsbruck: OKLC.

Antonellis, I., Bouras, C., & Poulopoulos, V. (2005, July 27-30). Game based learning for mobile users. In *Proceedings of the 6th International Conference* *on Computer Games: AI and Mobile Systems (CGAIMS 2005).* Lousville, Kentucky.

Becta ICT Research (2003). *Handheld computers (PDAs) in schools.* Department for education and skills.

Brimijoin, K. (2003). Using data to differentiate instruction. *Educational Leadership, 60*(5), 70-3.

Buyukkokten, O., Garcia-Molina, H., & Paepcke, A., & Winograd, T. (2000). Power browser: Efficient web browsing for PDAs. In *proceedings of the Conference on Human Factors in Computing Systems, CHIOO* (pp. 43-437).

Carver, C.A., Howard, R.A., & Lane, W.D. (1999). Enhancing student learning through hypermedia courseware and incorporation of student learning style. *IEEE Transactions on Education, 42*(1), 33-38.

Chabra, T., & Figueiredo, J. (2002). How to design and deploy handheld learning. Retrieved on December 8, 2006, from http://www.empoweringtechnologies.net/eLearning/eLearning_expov5_files/frame.htm

Deng, Y. C., Chang, H. Z., & Chang, B., & Liao, H. C., & Chiang, M. C., & Chan, T. W.(2005). "Ask the author": An academic conference supported system using wireless and mobile devices. In *Proceedings of the third IEEE WMTE2005* (pp. 29-33).

Della-Dora, D., & Blanchard, L. (1979). *Moving toward self-directed learning: Highlights of relevant research and promising practice.* Alexandria, VA: Association for Supervision and Curricular Development.

Dunn, R. (2000). Learning styles: Theory, research, and practice. *National Forum of Applied Educational Research Journal, 13*(1), 3-22.

Dye, A., Solstad, B.E., & Joe, A. K.(2003). Mobile education—A glance at the future. Retrieved on July 15, 2006, from http://www.nettskolen.com/forskning/mobile_education.pdf

Felder, R.M. (1995). Learning and teaching styles in foreign and second language education. *Foreign Language Annals, 28*(1), 21-31.

Felder, R.M. (1996). Matters of style. *ASEE Prism, 6*(4), 18-23.

Felder, R., Silverman, L. (1988). Learning and teaching styles in engineering education. *Engr. Education, 78*(7), 674-681.

Felder, R.M., & Soloman, B.A. (2005). *Index of Learning Styles Questionnaire.* Retrieved on July 14, 2006, from http://www.ncsu.edu/felder-public/ILSdir/ilsweb.html

Fischer, G., & Konomi, S.(2005). Innovative Media in Support of Distributed Intelligence and Lifelong Learning. In *Proceedings of the 3rd IEEE WMTE2005* (pp. 3-10).

Gay, G., & Reiger, R. & Bennington, T. (2001). Using mobile computing to enhance field study. In Miyake, N., Hall R, & Koschmann, T. (Eds.), *Carrying the conversation forward.* Mahwah, NJ: Erlbaum 507528.

Gay, G., Stefanone, M., & Grace-Martin, M., & Hembrooke, H. (2002). The effect of wireless computing in collaborative learning environments. *International Journal of Human-Computer Interaction, 13*(2), 257-276.

Harris, P. (2001). Goin Mobile. Retrieved on July 3, 2006, from http://www.learningcircuits.org/2001/jul2001/harris.html

Heoyoung, K. (2002). *Development of wire and wireless learning system based on XML.* Dissertation, Keimyung University.

Hoppe, H.U., Joiner, R., & Milrad, M., & Sharples, M. (2003) Guest editorial: Wireless and mobile technology in rducation. *Journal of Computer Assisted Learning, 19*(3), 255-259.

Kawachiya, K., & Ishikawa, H. (1999). Improving Web interaction on Small Displays. In *Proceedings of 8th international WWW Conference* (pp. 51-59).

Keefe, J. W. (1982). Assessing student learning style: An overview. In J. W. Keefe (Ed.), *Student learning style and brain behavior* (pp. 43-53). Reston. VA: National association of Secondary School Principals.

Ketamo, H. (2002). mLearning for kindergarten's mathematics teaching. *IEEE International Workshop on Wireless and Mobile Technologies in Education*. Retrieved on March 14, 2005, from http://csdl.computer.org/comp/proceedings/wmte/2002/1706/00 1706toc.htm

Kolb, D. (1984). *Experiential learning: Experience as the source of learning and development.* Englewood Cliffs, NJ: Prentice-Hall.

Kristiansen, T. (2001). *M-learning. Experiences from the use of WAP as a supplement in learning.* Oslo: Fornebu Knowation.

Kruzich, J.M., Friesen, B.J., & Van Soest, D.(1986). Assessment of student and faculty learning styles: Research and application. *Journal of Social Work Education, 22*(3), 22-30.

Landers, P. (2005). From e-learning to m-learning. *e-Learning and m-Learning.* Retrieved on March 14, 2006, from http://learning.ericsson.net/mlearning2/project_one/index.html

Laura, N., Lonsdale, P., & Vavoula, G., & Sharples, M. (2006). *Literature Review in Mobile Technologies and Learning.* FUTERLAB SERIES. Retrieved on July 14, 2006, from http://www.futurelab.org.uk/research/lit_reviews.htm

Lonsdale, P., Baber, C., Sharples, M., Byrne, W., Arvanitis, T., Brundell, P., & Beale, H. (2004). Context-awareness for MOBIlearn: Creating an engaging learning experience in an art museum. Proceedings of MLEARN2004, Bracciano, Rome: LSDA.

Lundin, J., Nulden, U., & Persson, L. M. (2001). *MobilLearn: Competence development for Nomads.* CHI 2001, 7-8.

Perry, D. (2003). Handheld computers (PDAs) in schools. Becta ICT Research Report, March 2003.

Sakamura, K., & Koshizuka, N. (2005). Ubiquitous computing technologies for ubiquitous learning. In *Proceedings of the third IEEE WMTE2005* (pp. 11-18).

Shindo, K., Koshizuka, N., & Sakamura, K. (2003). Ubiquitous information system for digital museum sing smart cards. In *Proceedings SSGRR*.

Schwabe, G., & Goth, C. (2005). Mobile learning with a mobile game: Design and motivational effects. *Journal of Computer Assisted Learning, 21*(3), 204-216.

Schön, D. A. (1983). *The reflective practitioner.* NY: Basic Books.

Sharples, M., Corlett, D., & Westmancott, O. (2002). The design and implementation of a mobile learning resource. *Personal and Ubiquitous Computing, 6*, 220-234.

Shotsberger, P. G., & Vetter, R. (2000). How mobile wireless technologies will changes web-based instruction and training. *Educational Technology, 40*(5), 49-52.

Stevens, A. (1999). *Tracking, curriculum differentiation, and the student experience in english (high schools, language arts).* State University Of New York at Albany**:** 1053.

Theroux, P.(2004). *Enhance learning with technology.* Retrieved on July 15, 2006, from http://www.enhancelearning.ca/

Tomlinson, C. A. & National Association for Gifted Children (U.S.) (2004). *Differentiation for gifted and talented students.* Thousand Oaks, CA: Corwin Press.

Vahey, P., & Crawford, V. (2005). *Palm Education Pioneers Program Final Evaluation Report.* Retrieved on March 14, 2006, from http://www.palmgrants.sri.com

KEY TERMS

Adaptive Mobile Learning Management System: The system managing the whole learning process and progress of a learner and containing the function to make learning adaptive to learning styles.

Learning Management System: A software system designed to facilitate teachers in the management of online educational courses for their students. These services generally include access control, provision of e-learning content, communication tools, and administration of user groups.

Learning Style: Described as general characteristics showing individual differences in the intrinsic procedures of information processing.

Mobile Device: Any portable device used to access the Internet. For example, PDA, cellular phone, Tablet PC and so on.

Mobile Learning: A form of learning where mobile computing is combined with e-learning and as a form of teaching and learning delivered through mobile devices such as mobile phones, PDAs, smart phones, tablet PCs, and so on.

Mobile Learning Environment: Provide students and teachers with the opportunity to obtain any and all class-related material on a Palm handheld computer through the network-enabled WAP Gateway.

Wireless Internet: Wireless Internet grants access to the World Wide Web or Internet e-mail via wireless networks.

Chapter XIX
Adaptive Interfaces in Mobile Environments:
An Approach Based on Mobile Agents

Nikola Mitrovic
University of Zaragoza, Spain

Eduardo Mena
University of Zaragoza, Spain

Jose Alberto Royo
University of Zaragoza, Spain

ABSTRACT

Mobility for graphical user interfaces (GUIs) is a challenging problem, as different GUIs need to be constructed for different device capabilities and changing context, preferences and users' locations. GUI developers frequently create multiple user interface versions for different devices. The solution lies in using a single, abstract, user interface description that is used later to automatically generate user interfaces for different devices. Various techniques are proposed to adapt GUIs from an abstract specification to a concrete interface. Design-time techniques have the possibility of creating better performing GUIs but, in contrast to run-time techniques, lack flexibility and mobility. Run-time techniques' mobility and autonomy can be significantly improved by using mobile agent technology and an indirect GUI generation paradigm. Using indirect generation enables analysis of computer-human interaction and application of artificial intelligence techniques to be made at run-time, increasing GUIs' performance and usability.

INTRODUCTION

Mobile computing is an increasingly important topic in today's computational environment, because the demand for ubiquitous access to information is constantly increasing. Furthermore, users want to increase their efficiency and process information when using mobile equipment. To support this demand, software applications face a number of challenges. One of the important challenges in mobile computation is user interaction.

The importance of user interfaces (UI) comes from the fact that UIs represent the first line of interaction between a user and a computer. A user's ability to execute a required task and his efficiency are directly impacted by the user interface.

In the past, user interfaces have been developed mostly for a specific device, for example a specific PDA variant or a work station. Such an interface was usually designed for a single platform. This was done in conjunction with specialized user interface libraries that were defined for a specific platform or programming language. For example, if we assume an application that is developed for Windows and UNIX platforms, the user interface for the windows platform would be designed and developed separately from the UNIX user interface. The cost and effort required for such a development are obviously high; such an approach frequently leads to other problems, for example GUI implementations on one (or more) platform(s) being at different levels of development due to the lack of resources required to maintain the same GUI version on multiple platforms.

In the mobile environment GUIs face additional challenges: a user could be using an application on a mobile phone and could require the same application on his PDA or WebTV. In addition, the user could be moving and requiring an application to move with him. For example: while in the car, a user could read his e-mail using car's on-board computer; when he steps out of the car he could prefer to continue working on his PDA until he gets to the office, where a desktop PC could be his preferred equipment to continue working.

Mobile devices have different capabilities and requirements: different processing power, screen size, supported colors, sound functionalities, keyboard, and so forth. In addition, mobile devices use an ever-increasing number of different hardware and OS solutions, and frequently rely on batteries for operation. Mobile applications use wireless networks; wireless networks are not stable, have limited capacity and performance, and are expensive (e.g., 3G networks).

In addition to this, application interface and functionality may change depending on a user's context. For example, a music player application should mute if the user is indoors and should turn on when outdoors. Furthermore, a user could prefer the speaker to be on a louder setting when in the car. These requirements could be either a user's preferences or rules associated with a particular location where the user is.

To meet such challenges, researchers in the user interface area have adopted a common approach—user interface abstraction. To be presented on a concrete platform, abstracted user interfaces are transformed and rendered to meet a concrete platform's requirements. This approach provides a single user interface definition that is later transformed to the target device's user interface. The abstraction level in such an abstract user interface definition varies. Some abstract user interface notations offer very abstracted descriptions of user interfaces, while others are more linked to specific user interface concepts, for example window-based user interfaces. An abstract user interface definition is usually delivered in XML (W3C, 2000) notation, which enables efficient processing and data exchange between multiple platforms. Some notations describe the user interface at a high level, for example, a button is required; others allow sophisticated definitions of constraints and additional parameters, such as requiring the button for some (specific) device(s) only.

This chapter presents different approaches to adapting user interfaces to devices, with specific interest focused on enabling architectures that adapt to users' preferences and contexts. We discuss difficulties with mobile user interface generation for wireless devices. Finally, we present an approach for user interface adaptation based on mobile agents and examine sample usage of such an approach. With this example we show a flexible and mobile generation of user interfaces in a wireless environment that allows monitoring of user interaction and application of knowledge.

BACKGROUND

User Interface Abstraction

Abstraction of user interfaces adds flexibility when generating user interfaces. It provides a single and comprehensive description of the user interface. It is a rich layer of information that describes a user's interaction with the computer. Such abstraction is formalized by using an abstract user interface definition language (Stottner, 2001) or task models (Limbourg & Vanderdonckt, 2003). This information can then be used to generate a user interface that meets a concrete platform's limitations and requirements. Such an approach develops a single set of information to support all variants of the user interface that should be created for different devices.

The user interface is typically abstracted through the use of design models (e.g., task models) or by using an XML-based abstract user definition language (Luyten & Coninx, 2001; Mitrovic & Mena, 2002; Molina, Belenguer & Pastor, 2003; Stottner, 2001). Task models (Limbourg & Vanderdonckt, 2003) provide information that is focused on tasks. Some of the task models can describe multi-modal tasks for different types of devices (Paterno & Santoro, 2002). However, task models do not necessarily specify the exact presentation of a user interface. On the other side, XML-based abstract user interface descriptions describe the user interface‘s presentation and constraints. Many different versions of such abstract user interface definition languages exist.

Such approaches' abstraction levels vary: some approaches include information specific to a device type (e.g., a mobile phone, or a specific mobile phone model), some are more generic and do not consider the specifics of any device type. In addition, user interface abstraction can differ conceptually—some models can define any kind of interaction (e.g., via voice or specialized interfaces), while some are more linked to specific concepts (e.g., window-based user interfaces). Examples of such abstraction languages include XUL, UIML, XIML and XForms (Stottner, 2001).

The ability to effectively adapt such user interface definitions to a concrete platform is a key factor in achieving mobile and efficient user interfaces. The resulting, concrete user interface must meet the specification and be functional on the target device. This requires that device capabilities and limitations be successfully addressed. In addition, user preferences and context frequently impact this adaptation.

Abstract User Interface Adaptation

User interface adaptation is a complex task, and includes not only adaptation to the specific device's capabilities, but also to the user. Mobile devices have different capabilities such as screen size, keyboard and support for particular user interface widgets, hardware platform or operating system. Adaptation to a user includes adaptation to the user's preferences and changing contexts, but sometimes includes factors such as previous knowledge or location.

Platforms may have exceptionally different user interface capabilities and requirements (see Figure 1). In many cases adapting a user interface simply as a per user interface specification is not sufficient. For example, a combo-box widget as specified in the user interface description may be available on a particular platform, or not; there could be a similar widget or this widget should be transformed into a set of different widgets. To address this and to maintain the user interfaces' plasticity (Thevenin & Coutaz, 1999) additional adaptation effort is required. User interface plasticity is a user interface's capacity to preserve usability regardless of variations in the system hardware specification or operating environment.

Adapting to users includes a wide range of considerations: users' preferences, context, location, ambient environment, and so forth. This is a more complex transformation than adaptation

Figure 1. Some of the target platforms for mobile applications (two versions of smart phone and a web TV)

to a device as it requires knowledge about the user. The key to this adaptation is the ability to understand the user's actions, location and other parameters of interest, and to apply different artificial intelligence techniques to such information. To achieve this goal it is very important to gather information surrounding the user, but also to gather information coming from the user himself. This information coming from the user could be in the form of predefined preferences (e.g., color scheme preference) but, more importantly, the data about interactions between the user and his computer are of great value for our purposes. By observing and analyzing the user's interaction, his corresponding interface could be modified according to the type of interaction (e.g., using stylus or keyboard), tasks previously executed by the user, or tasks executed by other users.

Adapting to device capabilities requires a program capable of processing or interpreting the abstract user interface definition in such a way that it can be rendered to a particular device. Two major approaches are used to achieve this:

- Using a design-time tool to create concrete user interfaces for the required platforms
- Using a run-time tool or process to render a concrete user interface

Design-time tools are closer than run-time techniques to traditional user interface development and frequently encompass initial modeling, analysis and later development of multiple user interfaces. Such approaches include design patterns (Seffah & Forbrig, 2002), task models (Limbourg & Vanderdonckt, 2003; Molina et al., 2003;Paterno & Santoro, 2002) or off-line analysis of user interaction (Pitkow & Pirolli, 1999). This approach provides good facilities for designing static user interfaces, but is challenged by mobility and unanticipated situations.

The design-time tools are usually specialized stand-alone tools used by the user interface or software developer. Run-time tools are usually more generic mechanisms built into the application or programming framework.

Run-time tools offer higher flexibility than design-time tools and reduce effort for user interface developers. However, they are less capable of producing fine-tuned user interfaces, which can be produced by design-time methods on a case-by-case basis. The run-time approach is better suited for applications that are accessed remotely or applications that require execution on mobile devices with varied GUI and processing capabilities.

In the following sections we will present the benefits and drawbacks of both design-time and run-time approaches, with a particular emphasis on techniques used to adapt user interfaces to various devices.

Adaptation to Devices

Adapting user interfaces to devices requires a program capable of processing an abstract user interface definition to a device's capabilities. Such a program must adapt the abstract specification to the target (concrete) user interface specification. The concrete user interface specification must represent user interface design well and remain functional on the target device.

Two major approaches are used for such adaptations:

- **Design-time:** This encompasses using a design-time tool to transform the abstract definition into a user interface adapted to a concrete user interface's specific capabilities. This approach offers flexibility for the user interface developer to fine-tune the user interface's generation; but requires additional effort and is challenged by mobility, as it is not portable to devices with different capabilities, which is a common situation in mobile computing.
- **Run-time:** This method uses a run-time tool to render an abstract user interface specification for the device on which such a GUI is needed. This method does not allow user interface developers to fine-tune user interfaces, but gives higher flexibility and mobility to developed user interfaces at a much lower effort than design-time approaches.

Design-Time Adaptation

The design-time adaptation approach is based on generating user interfaces at design-time, as opposed to the run-time approach. Typically, a user interface definition is created for an application. This definition is then transformed using a tool into a concrete user interface for a specific device make or model. Such transformations usually generate different program source code for different interfaces (e.g., Visual Basic, Java or others). Source code is then compiled for the required platform and then executed on the platform. The most common approach is to define a multi-platform task model (Paterno & Santoro, 2002) and then to generate different user interfaces from this model. User interfaces generated in this way can be executed only on a specific device and platform; in contrast, the run-time approach tends to be much more versatile with respect to device capabilities.

Molina et al investigated an approach to define user interface using models and then automatically generating different programs (program source code) for different platforms (Molina et al., 2003). Code generators in this work are created for a limited number of different programming languages/platforms; however, additional code generators can be added to accommodate additional languages or platforms.

A similar approach, using an abstract user interface definition, was developed by Microsoft Corporation (2005) for its Longhorn/Avalon platform—it defines a user interface using an abstract notation, and then programmers develop code for different devices. This centralizes the user interface design but still requires multiple code implementations for different devices. In addition, Microsoft's approach is supported only on Windows platforms.

In general, a design-time approach offers some flexibility for the user interface developer because the resulting user interface can be manually fine-tuned before it is executed on the platform. However, the effort and expertise required for inspecting and fine-tuning multiple GUI versions before compilation and execution can be time-consuming and costly. Some of the design-time approaches still require programmers to develop multiple code implementations to handle user interface rendering and interaction on different platforms.

Run-Time Adaptation

Run-time adaptation is performed by using a program to adapt user interface definition at the time of program execution. Several approaches are used for run-time adaptation:

- **Standalone adaptation:** A specialized program adapts an abstract user interface definition to a specific platform.
- **Client-server adaptation:** A client program communicates with the server program in order to generate and present the user interface.
- **Mobile agent adaptation:** Mobile agents within a mobile agent platform compose a mobile application that generates a mobile user interface.

In this section we will examine standalone and client-server adaptation; mobile agents and approaches based on them are detailed in the next section.

A standalone adaptation is delivered through a specialized program that adapts an abstract user interface definition to a specific platform. An example of such an adaptation is a Windows program that adapts an XML user interface definition to the Windows platform. Multiple implementations of adaptation programs can be developed for different platforms—for Java platform, Palm PDAs, and so forth. Luyten and Coninx (2001) developed a platform that utilizes an abstract user interface definition that is later rendered by multiple middleware software to various platforms. This approach can provide application functionality by a specialized proxy server, rather than by the mobile device itself. In this case data and program functionality are then handled using Web services. This could present a limitation in situations where a wireless connection is not available or is not performing well.

A similar approach was adopted by Microsoft Corporation (2002), with "mobile forms" which automatically render to platforms supported by Microsoft's underlying engine, but which offers only a limited set of UI widgets (X Org, 1984). In addition, Microsoft's approach is available only on platforms supporting Windows and Microsoft Mobile Forms. This also restricts the number of available (supported) hardware platforms.

Such an adaptation approach lacks the mobility and flexibility required for mobile applications. The user interface is derived from a single specification, but it is pre-loaded onto devices and each software update triggers a new software set-up. The applications and user interfaces are frequently locked for a single device make and model. Development effort is high as multiple programs should be developed and verified for multiple platforms. In addition, mobile devices may not be capable of executing application functionality and, in such cases, specialized designs have to be developed. This introduces additional complexity and makes software more difficult to develop, maintain, and support.

The MobiLife project (IST MobiLife, 2006) investigates the creation of user interfaces for mobile devices. This project studies the use of multimodal user interfaces defined in an XML-based notation with the use of Web services (Baillie et al., 2005). Among other areas, MobiLife focuses on context awareness, sharing and personalization (Kernchen et al., 2006; Salden et al., 2005). The MobiLife approach is based on a client-server architecture (Baillie et al., 2005) and is limited in terms of network mobility and autonomy: platform-specialized program implementations (clients) are used—network mobility is limited only to compatible platforms; autonomy of clients is restricted—autonomy depends on the availability of corresponding server components.

The client-server adaptation is based on the client-server computation model. A server platform works in conjunction with clients and usually provides computation for (less capable) clients. For example, a Web application that is capable of transforming abstract UI definitions to both HTML and WML follows a client-server adaptation model. Such an approach requires a specialized client-side program capable of interpreting server-side information. Thus, multiple client programs must be developed for different devices that introduce difficulties in maintaining client programs and their versions for a variety of devices. In addition, a level of anticipation is required in this approach in order to create constantly functional client programs. For example, if the user interface definition for a particular device is correctly specified, but requires more processing power (e.g., the minimal memory required for processing all UI elements exceeds the available memory for the specific mobile device), then it is anticipated that the client-program would fail unless this has been anticipated by the client-program developer at the time of development.

IBM (2002) developed the WebSphere Transcoding Publisher server, which could transform a user interface specification into HTML, WML and other formats. The program could also handle JavaScript and transformations to WmlScript. However, this tool does not provide true mobility and is designed for Web applications. A similar approach, aimed mostly at Web development, is present in the Java community under the name Java Server Faces (JSF) (Sun Microsystems, 2005). JSF is focused mainly on Web development and is server oriented. This framework provides a component-based framework for Web user interface development. In addition, this framework allows the specification and development of alternative renders—a rendering engine can potentially render a custom user interface definition to different platforms. For example, a render could transform a JSP or XUL UI definition to WML.

In the server-client model, a client-side program must be developed for all supported platforms, which introduces additional development, support, and management complexity, and could pose limitations when adapting to different platforms. This approach, however, provides limited mobility for the applications, as one application can be used from a number of different platforms. The client-server model relies on a good network connection, which is a limiting factor in costly, unstable or low-performing mobile networks.

Mobile Agent Adaptation

This adaptation approach is based on mobile agent technology (FIPA, 2002; Milojicic et al., 1998). The mobile agent technology eases automatic system adaptation to its execution environment. A mobile agent is a program that executes autonomously on a set of network hosts on behalf of an individual or organization. Software agents can easily adapt their behavior to different contexts. Mobile agents can bring computation wherever needed and minimize network traffic, especially in wireless networks (expensive, slow and unstable). In addition, mobile agents do not decrease the system's performance, even when communications are based on a wired network, as shown in (Mitrovic, Royo & Mena, 2005). Mobile agents can arrive at the users' device and show their GUIs to the users in order to interact with them. Mobile agents can be hosted by platforms that support different models of user interfaces or that have different processing capabilities. Because mobile agents are autonomous they can handle communication errors (unreachable hosts, etc.) by themselves. Also, they can move to the target device instead of accessing target devices remotely. For instance, agents can be sent to a home computer supporting Java and Swing, or they can play the role of a proxy server for a wireless device, such as mobile telephone or a Web terminal; in that case they should produce an adequate GUI—WML or HTML (Mitrovic & Mena, 2002). Such an adaptation is not limited to Web pages or mobile phones; other devices such as PDAs benefit from such architectures (Mitrovic, Royo & Mena, 2005).

Mobile agents' platforms (Bursell & Ugai, 1997; Grasshopper, 2000; Ilarri, Trillo & Mena, 2006) are usually based on Java due to its portability, but are not limited to any particular platform (Wang, Sørensen & Indal, 2003). Mobile agent platforms are being developed for both old and new platforms such as Java Micro Edition (J2ME) (Wang et al., 2003).

Mobile agents can incorporate various learning techniques and learn from past experiences (Mitrovic & Mena, 2003). This is particularly important in the mobile world, where the abilities to learn and adapt to contexts and users are some of the most important requirements.

The mobile agent approach brings higher mobility and flexibility of user interaction while reducing development complexity—only one version of a program is defined.

To summarize, user interface adaptation to different devices is difficult; approaches are challenged by true mobility, require multiple development efforts, and have various degrees of flexibility and transparency. Mobile agents eliminate the majority of these issues and provide a flexible platform for the development of adaptive and portable GUIs.

Adaptation to Users

Different approaches are used to adapt user interfaces to meet users' preferences and contexts. Similar to adaptation to devices, adaptation to users can be done at design-time or at run-time. The design-time techniques include design patterns (Seffah & Forbrig, 2002), analysis of UI usage (Pitkow & Pirolli, 1999) or usability tests followed by UI redesign. Run-time techniques are present mostly in the Web arena, where Web-page content is rearranged by users' preferences or the Website's context (Amazon Online bookshop, 2006). Run-time techniques are based mostly on statistical and data mining techniques (Zukerman & Albrecht, 2000), which apply various artificial intelligence techniques to interaction data.

An additional challenge for mobile applications and run-time adaptation is users' attention (Vertegaal, 2003). Users' attention is a limited resource that is required by many external stimuli, including software applications. It is important to prioritize and adapt requests coming from software applications toward user in order to most effectively utilize this resource. Therefore, user interaction timing and volume must be considered when designing attentive user interfaces. Monitoring users' attention through physical indicators such as eye movement, geographic location, or statistical data mining is crucial for prioritizing, adapting and designing interaction tasks (Vertegaal, 2003).

To be able to effectively adapt UIs to users at the run-time, it is crucial that systems provide facility for collecting information on computer-user interaction. Such facility enables later analysis and exchange of collected information, which could lead to changes in the application logic or user interface. Therefore, another requirement for successful UI adaptation to users is a facility for changing the user interface at run-time. Using this facility, or architecture, the user interface could be easily amended according to the collected information.

The majority of applications, providing such facilities, are not designed with interoperability in mind and focus only on the current application's ability to gather and analyze interaction information. Applications frequently cannot exchange interaction information, and applications programmed by different developers cannot use a common set of facilities. This leads to a multiplication of developments designed for just a single application or application vendor. However, some systems based on mobile agents can be used by multiple applications and different (independent) learning modules.

ADUS: AN APPROACH BASED ON MOBILE AGENTS

The following introduces ADUS, our proposal for indirect generation of adaptive and portable GUIs. ADUS—ADaptive User Interface System (Mitrovic et al., 2004) is a system based on mobile agent technology that helps with user interface generation and allows monitoring of user-computer interaction. ADUS is part of the ANTARCTICA system (Goñi, Illarramendi, Mena et al., 2001)—a multi-agent system that provides users with different wireless data services to enhance their mobile devices' capabilities.

ADUS has three main functions: (1) transparently adapting an abstract user interface definition to a concrete platform, (2) monitoring user interaction and communicating this information to other agents, and (3) communicating and collaborating with other agents in the ANTARCTICA platform. ADUS also performs a number of agent-based functions such as optimized network operation and collaboration with other agents. The ADUS system uses XUL (XUL Tutorial, 2002), an XML-based abstract user interface definition language.

Adaptive User Interface Generation in ADUS

In this section we present and discuss several approaches to generate adaptive user interface allowing the monitoring of the user behavior.

Option 1: The Visitor Agent Generates the GUI

The first approach is when the visitor agent arrives at the user's device it requests from the user agent the available resources, the user's preferences and the device's capabilities. Then the visitor agent creates the GUI by itself and interacts with the user directly.

This approach solves the generation of customized GUIs. However, it still has several problems:

- The user agent cannot monitor the user's behavior because the data provided to the GUI flow directly to the visitor agent.
- The user agent must trust the visitor agent to render a GUI according to the user's preferences and the device's capabilities. Visitor agents could ignore the user agent descriptions and try to show their own GUI (in that case, the type of GUI created by the visitor agent could not be executed on that device).
- All the visitor agents have to know how to process and apply the knowledge provided by the user agent (which implies that they all must know how to generate any kind of GUI).

Option 2: The User Agent Generates the GUI and Delegates Event Handling to the Visitor Agent

In this approach the visitor agent, after arriving at the user's device, provides the user agent with a specification of the needed GUI. Then the user agent generates a GUI according to the user's preferences, the device's capabilities and the visitor agent's requirements, and it delegates the GUI event handling to the visitor agent.

The user interface specification can be made in the Extensible User-interface Language (XUL) (XUL Tutorial, 2002). This interface definition can be later adapted by the user agent using XSL transformations to the required GUI representation language (HTML, WML, etc.). The XUL interpretation on Java-enabled platforms is interpreted by a Java XUL platform that renders XUL using standard AWT and Swing widgets.

The advantages of this approach are:

- The user agent guarantees that the visitor agents' GUIs will be generated correctly (according to the user's preferences and the device's capabilities) if they are specified in XUL.
- Visitor agents do not need to know how to generate GUIs in different devices.
- The user agent can deny permission to generate GUIs to all visitor agents (Mitrovic & Arronategui, 2002) in order to avoid direct GUI generation.

However, following this approach, the user agent cannot monitor the user's behavior because GUI events are handled directly by visitor agents. Therefore the user agent must trust the visitor agent to get the information about its interaction with the user.

Option 3: An Intermediate Agent Generates the GUI and Handles the Events

In this approach, first the visitor agent sends its XUL specification of the GUI to the user agent, then the user agent generates the GUI and handles all the events (it receives data from the user), and finally, it sends the user data back to the visitor agent.

This approach has all the advantages of the approaches presented above. Furthermore, it allows the user agent to monitor the user's behavior easily and efficiently as it handles the GUI events. Although this approach is interesting, its implementation faces a problem: the user agent must attend the different services executed on the user's device, and some tasks, such as GUI generation, could overload it. Therefore, a better approach is for the user agent to delegate the generation of adaptive GUIs to a specialized agent (the ADUS agent). Thus, the distribution of the service execution across the three agents (the ADUS agent, the user agent and the visitor agent) allows us to balance the system's load.

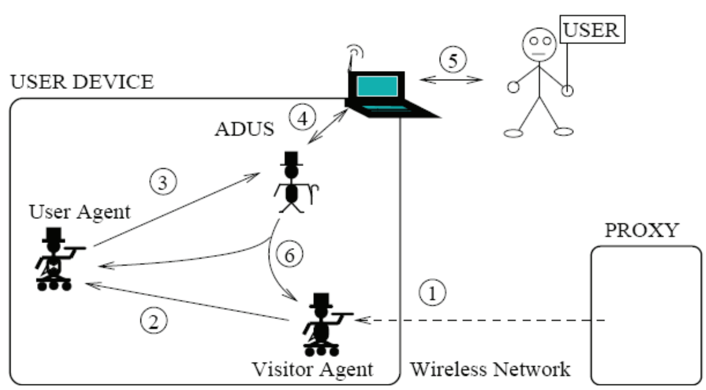
Indirect Generation of GUIs

This section describes in more detail the architecture needed to efficiently generate adaptive GUIs. To illustrate this process we use an example application. As shown in Figure 2, the system contains the following agents:

- **The visitor agent:** This is a mobile agent that brings a service requested by the user to the user's device. This agent can generate an XUL specification of the GUIs that it needs to interact with the user. Such XUL specifications are sent to the user agent on the user's device.
- **The user agent:** This is a highly specialized personalization agent that is in charge of storing as much information as possible about the user and his computer. For example, it knows the user's look and feel preferences and the GUI preferred by the user or imposed by the user's device or the operating system. This agent's main goals are: (1) to proxy the generation of user interfaces, (2) to help the user to use the visitor agent's services, (3) to modify the visitor agent's GUI specification according to the user's preferences, (4) to create an ADUS agent initialized with the static GUI features such as the user's device capabilities, and (5) to monitor user interactions by receiving such information from the ADUS agent.
- **The ADUS agent:** This agent's main activities are: (1) to adapt the user interface to the user's preferences and device capabilities, following the user agent's suggestions; (2) to generate GUIs for different devices according to the XUL specification; and (3) to handle GUI events and communicate them to the visitor and user agents (allowing the user agent to monitor the user's interaction). There is one ADUS agent per visitor agent.

The following describes the synchronization of the previously mentioned agents. As an example, we use a simple currency converter application that converts between currencies per a user's requests, and displays the results of the conversion. This application is executed by mobile agents that travel to the user's device when requested by the user. The main steps are (see Figure 2):

Figure 2. Indirect generation of GUIs



1. The visitor agent travels to the user's device. This step is only for approaches that are based on mobile agents. For example, it is equivalent to the call of a local application (in client-server architecture).

   The visitor agent requests the generation of its GUI. In this step the visitor agent sends the XUL description of its GUI to the user agent. In Figure 3 we show the XUL definition of the GUI for the currency converter service.
2. The user agent processes the GUI specification, transforms the GUI description to adapt it to the user's preferences, and creates the corresponding ADUS agent initialized with: (1) the transformed XUL description of the GUI to generate, and (2) the static information for the GUI such as the device's capabilities, screen resolution, representation language of the user's device (WML, HTML, Java Swing, etc.), among other information.
3. The ADUS agent generates the GUI according to the information provided by the user agent (static GUI information and specific information for this service). The ADUS agent can map any XUL description into GUIs for devices with different features; for example, a WAP device or a laptop with a Java GUI.

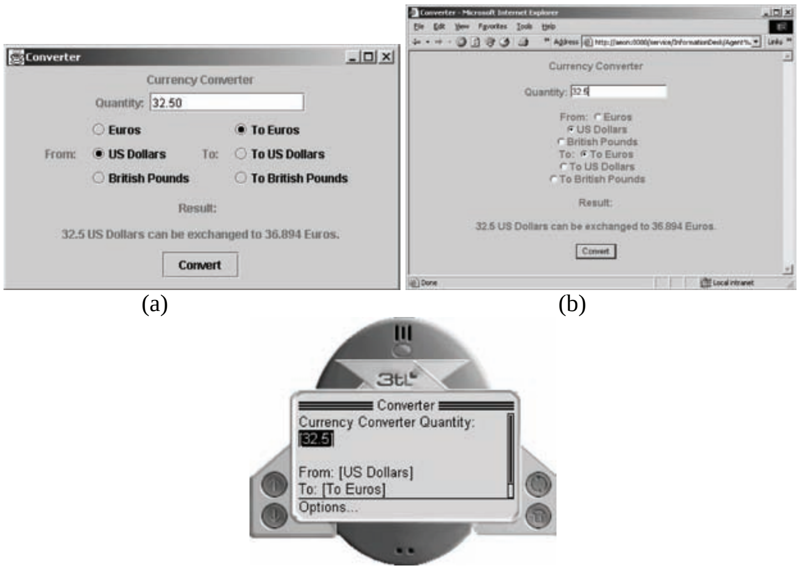
In the example, if the converter application were executed on a device with Java Swing capabilities (e.g., a home PC or laptop) the ADUS agent would generate a Swing GUI (see Figure 4a). When the converter application is executed on a Web terminal without applet support, the generated GUI would be based on HTML, as shown in Figure 4b. Finally, if it is executed on a WAP mobile phone, then the GUI is based on WML, as shown in Figure 4c. We point out that this functionality of the ADUS agent works for any XUL specification. The ADUS agent could be extended with mappings to other kinds of GUI languages, such as Macromedia Flash or J2ME.

4. The user interacts with the GUI by looking at the information presented on the screen device and by using the device's peripherals (keyboard, mouse, buttons, etc.) to enter data or select among different options.
5. The ADUS agent handles and propagates the GUI events. User actions trigger GUI events that are captured by the ADUS agent. This information is sent to: (1) the visitor agent, which reacts to user actions according to the service that it executes, perhaps by generating a new GUI (step 2), and (2) the user agent, which can store and analyze the information provided by the user in order to reuse it in future service executions. One of the advantages of the presented architecture is that both messages can be sent concurrently, so the load is balanced.

Figure 3. XUL definition for the currency converter service

```
...
<vbox>  <hbox>
  <label control="lblQty" value="Quantity:"/>
  <textbox value="0.00" id="Qty" size="20"/>
 </hbox>  </vbox>
 ...
<-- label that will be used for the Output -->
<box>
 <label control="lblOutput" value="Result:  "/>
</box>
<box>
 <label id="Output" control="Output" value=""/>
</box>
<!-- adding button -->
<box>
  <button id="Convert" label="Convert"  oncommand="convert()"/>
</box>
...
```

Figure 4. Currency converter service (swing, HTML and WAP transformation)



Finally, we would like to stress the relevance to the user agent of monitoring the user's interaction with visitor agents. By knowing the user's reactions and data entered on those services, the user agent can store such data locally and apply different artificial intelligence techniques to extract knowledge about the user's behavior. In the previous example, the next time that the currency converter service executes on the user's device, the user agent could select (in the visitor agent's XUL specification) U.S. dollars and Euros as the initial and target currencies, respectively, because that was the user's selection during the previous execution of that service. Even if the user now selects another configuration, the user agent could learn and improve its behavior for the next time. In addition, the user interface could be rearranged to meet the user's preferences, for example the "convert" button could be in a different part of the screen. Thus, the customization of GUIs can become really useful for the user, as the user agent can monitor, store and analyze his interactions with all the GUIs/applications. In this case, mobile agents would learn from interaction data of other applications and from interaction data collected by other users (Mitrovic & Mena, 2003).

FUTURE TRENDS

In our increasingly mobile world, human-computer interaction has become one of the most important topics. Mobile user interface systems are more and more required to be context and location-aware, and we should improve interfaces by applying knowledge-based techniques. The mobility of systems providing a mobile user interface is an important factor, and mobile agents are one of the promising technologies in this field. Significant improvement in mobile agent platforms' scalability and performance is underway (Ilarri et al., 2006).

From the adaptation perspective, trends are likely to be the further standardization of UI definition languages. Currently many different variants of XML-based user interface languages exist, which makes collective, community efforts less efficient than they would be if only one such language existed. Industry bodies such as the WWW Consortium have already begun proposing standard approaches, such as XForms (WWW Consortium, XForms, 2003). Improvements in mobile user interfaces' plasticity are a challenge yet to be fully addressed by researchers.

Consolidating mobile devices' capabilities and constantly increasing the performance of devices and infrastructure are two of the forces driving mobile systems and mobile user interface systems. Because processing power and different capabilities are two of the key issues for mobile user interfaces, this will positively affect both research efforts and developments. Although mobile devices are constantly being improved, certain device properties, such as screen size, will require methods that allow adaptation to users' devices.

In the agent-based user interfaces area, future trends are in increased interoperability across different multi-agent platforms. This will improve agent-based user interface systems' cross-platform mobility.

CONCLUSION

Generation of adaptive user interfaces in mobile environments is a complex task that faces many challenges. Mobile devices have different capabilities, work in different contexts, and have different kinds of users. Users require adaptation to their context and preferences. Mobile networks remain an expensive and unreliable medium, which presents additional difficulties. Many different approaches are used to solve user interfaces' problems in a mobile environment, from design-time considerations to various run-time solutions.

Design-time solutions are inclined more toward traditional approaches and are based more on multi-modal user interface modeling and code developments for different mobile devices than are run-time solutions. These approaches do not offer mobility and could suffer from mobile devices' low processing capability. Design-time approaches usually allow fine tuning of the user interface code (before deployment), which requires an additional development effort and expertise in the specific mobile system.

Run-time solutions offer a degree of mobility and higher flexibility than design-time solutions. The quality of the automatically generated interfaces is lower than in design-time approaches because of the lack of fine-tuning the user interface code before deployment. Run-time systems based on a client-server model give a degree of mobility but are limited by the requirement for mobile networks' constant availability. In addition, some kind of client software is required, which represents an additional effort.

However, mobile agents as a run-time approach provide good mobility for user interfaces and increased flexibility over other approaches. Mobile user interfaces based on mobile agents can transparently (to developers and users) adapt user interfaces to devices "on the fly." In addition, agents can distribute the processing load appropriately so that devices with low processing capabilities can still execute complex applications. Systems based on mobile agents can monitor and analyze computer-user interaction, and share such information between different program instances, users or systems. Such information can then be used to improve application logic and to create additional user interface adaptations. As mobile agents are fully mobile software entities, they enable mobile-agent based software and user interfaces to follow users wherever required.

Topics yet to be fully explored by researchers in the agent-based user interfaces area include inter-platform interoperability. Providing higher plasticity than is currently seen in mobile user interfaces remains a challenging task for all mobile user interfaces.

REFERENCES

Amazon online bookshop. (2006). Retrieved from Amazon: www.amazon.com

Baillie, L., Schatz, R., Simon, R., Anegg, H., Wegscheider, F., Niklfeld, G., & Gassner, A. (2005). Designing Mona: User interactions with multimodal mobile applications. *HCI International 2005, 11th International Conference on Human-Computer Interaction* (pp. 22-27).

Bursell, M., & Ugai, T. (1997). *Comparison of autonomous mobile agent technologies.* Retrieved from APM: www.ansa.co.uk/ANSATech/97/Primary/198901.pdf

FIPA, Foundation for Intelligent Physical Agents. (2002). Retrieved from www.fipa.org

Goñi, A., Illarramendi, A., Mena, E., Villate, Y., & Rodriguez, J. (2001). ANTARCTICA: A Multiagent System for Internet Data Services in a Wireless Computing Framework. *NSF Workshop on an Infrastructure for Mobile and Wireless Systems* Vol. 2538 (pp. 119-135). Germany: Springer Verlag Lecture Notes in Computer Science LNCS.

Grasshopper. (2002). Retrieved from http://www.cordis.europa.eu/infowin/acts/analysys/products/thematic/agents/ch4/ch4.htm

IBM. (2002). Retrieved from IBM WebSphere Transcoding Publisher: http://www-3.ibm.com/software/webservers/transcoding/

Ilarri, S., Trillo, R., & Mena, E. (2006, June). SPRINGS: A scalable platform for highly mobile agents in distributed computing environments. In *proceedings of the 4th International WoWMoM 2006 workshop on Mobile Distributed Computing (MDC'06).* (pp. 633-637). New York: IEEE Computer Society.

IST MobiLife. (2006). Retrieved from http://www.ist-mobilife.org/: IST MobiLife

Kernchen, R., Bonnefoy, D., Battestini, A., Mrohs, B., Wagner, M., & Klemettinen, M. (2006). Context-Awareness in MobiLife. *15th IST Mobile Summit, Mykonos, Greece*.

Limbourg, Q., & Vanderdonckt, J. (2003). Comparing task models for user interface design. *The Handbook of Task Analysis for Human-Computer Interaction*. Lawrence Erlbaum Associates.

Luyten, K., & Coninx, K. (2001). An XML runtime user interface description language for mobile computing devices. In Johnson, C. (Ed.), *Design, Specification and Verification of Interactive Systems - DSV-IS 2001* (pp 1-15), Germany: Springer Verlag Lecture Notes in Computer Science.

Microsoft Corporation. (2002). *Creating mobile web applications with mobile web forms in visual studio.NET.* Creating Mobile Web Applications with Mobile Web Forms in Visual Studio.NET. Retrieved from http://msdn.microsoft.com/vstudio/technical/articles/mobilewebforms.asp

Microsoft Corporation. (2005). *Longhorn Developer Center Home: Avalon*. Avalon. Retrieved from http://msdn.microsoft.com/Longhorn/understanding/pillars/avalon/default.aspx

Milojicic, D., Breugst, M., Busse, I., Campbell, J., Covaci, S., Friedman, B., Kosaka, K., Lange, D., Ono, K., Oshima, M., Tham, C., Virdghagriswaran, S., & White, J. (1998). MASIF, The OMG Mobile Agent System Interoperability Facility. *Mobile Agents,* Vol. 1477 (pp 50-68). Germany: Springer Verlag Lecture Notes in Artificial Intelligence - LNAI.

Mitrovic, N., & Arronategui, U. (2002). Mobile agent security using proxy-agents and trusted domains. *Second International Workshop on Security of Mobile Multiagent Systems (SEMAS 2002)* (pp. 81-83). Germany: German AI Research Center DFKI.

Mitrovic, N., & Mena, E. (2002, June). Adaptive user interface for mobile devices. In P. Forbrig, Q. Limbourg, B. Urban & J. Vanderdonckt (Eds.), *Interactive systems. Design, specification, and verification. 9th International Workshop DSV-IS 2002*, vol. 2545 (pp. 47-61). Germany: Springer Verlag Lecture Notes in Computer Science LNCS.

Mitrovic, N., & Mena, E. (2003). Improving user interface usability using mobile agents. In *Interactive Systems. Design, Specification, and Verification. 10th International Workshop DSV-IS 2003*, vol. 2844 (pp. 273-288). Madeira, Portugal: Springer Verlag Lecture Notes in Computer Science LNCS.

Mitrovic, N., Royo, J. A., & Mena, E. (2004). ADUS: Indirect Generation of User interfaces on Wireless Devices. *Fifteenth International Workshop on Database and Expert Systems Applications (DEXA)* (pp. 662-666). IEEE Computer Society.

Mitrovic, N., Royo, J.A., & Mena, E. (2005, September). Adaptive user interfaces based on mobile agents: Monitoring the behavior of users in a wireless environment. *Symposium on Ubiquitous Computing and Ambient Intelligence (UCAmI 2005)* (pp. 371-378). Spain: Thomson-Parninfo.

Molina, P. J., Belenguer, J., & Pastor, O. (2003, July). Describing just-UI concepts using a task notation. In J. Joaquim, N. Jardim Nunes & J. Falcao e Cunha (Eds.), *Interactive systems. Design, specification, and verification. 10th International Workshop DSV-IS 200*, vol. 2844 (pp. 218-230). Germany: Springer Verlag Lecture Notes in Computer Science LNCS.

Paterno, F., & Santoro, C. (2002). One model, many interfaces. In C. Kolski & J. Vanderdonckt (Eds.), *Fourth International Conference on Computer-Aided Design of User Interfaces* (pp. 143-154). Dordrecht: Kluwer Academics.

Pitkow, J., & Pirolli, P. (1999). Mining longest repeatable subsequences to predict world wide web surfing. In *proceedings of the 2nd Usenix Symposium on Internet Technologies and Systems (USITS)*, (pp. 139-150). USITS.

Salden, A., Poortinga, R., Bouzid, M., Picault, J., Droegehorn, O., Sutterer, M., Kernchen, R., Rack, C., Radziszewski, M., & Nurmi, P. (2005). Contextual personalization of a mobile multimodal application. *International Conference on Internet Computing (ICOMP)* (pp. 650-665). Las Vegas, USA: International Conference on Internet Computing (ICOMP).

Seffah, A., & Forbrig, P. (2002). Multiple user interfaces: Towards a task-driven and patterns-oriented design model. In *proceedings of the 9th International Workshop on Design, Specification and Verification of Interactive Systems (DSV-IS)* Vol. 2545 (pp. 118-133). Germany: Springer Verlag Lecture Notes in Computer Science LNCS.

Stottner, H. (2001). *A platform-independent user interface description language* (16). Linz, Austria: Institute for Practical Computer Science, Johannes Kepler University.

Sun Microsystems. (2005). *Java Server Faces*. Retrieved from JSF: http://java.sun.com/j2ee/javaserverfaces/

Thevenin, D., & Coutaz, J. (1999). Plasticity of user interfaces: Framework and research agenda. In *Proceedings of IFIP TC 13 International Conference on Human-Computer Interaction INTERACT'99, Vol. 1* (pp. 110-117). IOS Press.

Vertegaal, R. (2003). Attentive user interfaces. *Communications of the ACM, 46*(3), 30-33

W3C. (2000). Retrieved from Extensible Markup Notation: http://www.w3.org/TR/REC-xml

Wang, A.I., Sørensen, C.F., & Indal, E. (2003). A mobile agent architecture for heterogeneous devices. *3rd IASTED International Conference on Wireless and Optical Communications (WOC 2003, ISBN 0-88986-374-1* (pp. 383-386). Norway: ACTA Press.

WWW Consortium, XForms. (2003). Retrieved from X Forms: www.xforms.org

X Org. (1984). *X Windows*. Retrieved from X Org: www.x.org

XUL Tutorial. (2002). Retrieved from XUL: http://www.xulplanet.com/tutorials/xultu/

Zukerman, I., & Albrecht, D. (2000). Predictive statistical models for user modeling. In A. Kobsa (Ed.), *User Modeling and User Adapted Interaction (UMUAI)*, Vol. 11, (pp. 5-18). *The Journal of Personalization Research. Tenth Aniversary Special Issue*. The Netherlands: Kluwer Academic Publishers.

NOTE

This work is supported by the CYCYT project TIN2004-07999-C02-02.

KEY TERMS

Abstract User Interface Definition: Platform-independent and technology-neutral description of the user interface.

Design-Time User Interface Adaptation: Manual adaptation of the user interface by a designer, analyst or software developer.

Indirect User Interface Generation: A method in which the mobile agent requiring user interaction does not create a user interface directly, but passes the user interface definition to another agent (specialized for the user and his mobile device) that creates the above mentioned user interface and acts as an intermediary between the user and the agent that requires interaction with the user.

Mobile Agent: A program that executes autonomously on a set of network hosts on behalf of an individual or organization. One of the key features of such agents is mobility.

Multi-Agent System: A system that allows concurrent operation and communication of multiple (mobile) agents.

Run-Time User Interface Adaptation: Automatic adaptation of the user interface by a program during its execution.

Transparency: Automatic adaptation to specific conditions or circumstances without implicit or explicit intervention from the user, user interface designer or software developer.

User Interface Plasticity: A user interface's capacity to preserve usability regardless of variations in systems' hardware specification or operating environment.

Chapter XX
Intelligent User Interfaces for Mobile Computing

Michael J. O'Grady
University College Dublin, Ireland

Gregory M.P. O'Hare
University College Dublin, Ireland

ABSTRACT

In this chapter, the practical issue of realizing a necessary intelligence quotient for conceiving intelligent user interfaces (IUIs) on mobile devices is considered. Mobile computing scenarios differ radically from the normal fixed workstation environment that most people are familiar with. It is in this dynamicity and complexity that the key motivations for realizing IUIs on mobile devices may be found. Thus, the chapter initially motivates the need for the deployment of IUIs in mobile contexts by reflecting on the archetypical elements that comprise the average mobile user's situation or context. A number of broad issues pertaining to the deployment of AI techniques on mobile devices are considered before a practical realisation of this objective through the intelligent agent paradigm is presented. It is the authors hope that a mature understanding of the mobile computing usage scenario, augmented with key insights into the practical deployment of AI in mobile scenarios, will aid software engineers and HCI professionals alike in the successful utilisation of intelligent techniques for a new generation of mobile services.

INTRODUCTION

Mobile computing is one of the dominant computing usage paradigms at present and encapsulates a number of contrasting visions of how best the paradigm should be realized. Ubiquitous computing (Weiser, 1991) envisages a world populated with artefacts augmented with embedded computational technologies, all linked by transparent high-speed networks, and accessible in a seamless anytime, anywhere basis. Wearable computing (Rhodes, Minar, & Weaver, 1999) advocates a world where people carry the necessary computational artefacts about their actual person.

Somewhere in between these two extremes lies the average mobile user, equipped with a PDA or mobile phone, and seeking to access both popular and highly specialized services as they go about their daily routine.

Though the growth of mobile computing usage has been phenomenal, and significant markets exists for providers of innovative services, there still exist a formidable number of obstacles that must be surpassed before software development processes for mobile services becomes as mature as current software development practices. It is often forgotten in the rush to exploit the potential of mobile computing that it is radically different from the classic desktop situation; and that this has serious implications for the design and engineering process. The dynamic nature of the mobile user, together with the variety and complexity of the environments in which they operate, provides unprecedented challenges for software engineers as the principles and methodologies that have been refined over years do not necessarily apply, at least in their totality, in mobile computing scenarios.

How to improve the mobile user's experience remains an open question. One approach concerns the notion of an application autonomously adapting to the prevailing situation or context in which end-users find themselves. A second approach concerns the incorporation of intelligent techniques into the application. In principle, such techniques could be used for diverse purposes, however, intelligent user interfaces (IUIs) represent one practical example where such techniques could be usefully deployed. Thus the objective of this chapter is to consider how the necessary intelligence can be effectively realized such that software designers can realistically consider the deployment of IUIs in mobile applications and services.

BACKGROUND

Research in IUIs has been ongoing for quite some time, and was originally motivated by problems that were arising in standard software application usage. Examples of these problems include information overflow, real-time cognitive overload, and difficulties in aiding end-users to interact with complex systems (Höök, 2000). These problems were perceived as being a by-product of direct-manipulation style interfaces. Thus, the concept of the application or user interface adapting to circumstances as they arose was conceived and the terms "adaptive" or "intelligent" user interfaces are frequently encountered in the literature. How to effectively realize interfaces endowed with such attributes is a crucial question and a number of proposals have been put forward. For example, the use of machine learning techniques has been proposed (Langley, 1997) as has the deployment of mobile agents (Mitrovic, Royo, & Mena, 2005).

In general, incorporating adaptability and intelligence enables applications to make considerable changes for personalization and customization preferences as defined by the user and the content being adapted (O'Connor & Wade, 2006). Though significant benefits can accrue from such an approach, there is a subtle issue that needs to be considered. If an application is functioning according to explicit user defined preferences it is functioning in a manner that is as the user expects and understands. However, should the system autonomously or intelligently adapt its services based on some pertinent aspect of the observed behavior of the user, or indeed, based on some other cue, responsibility for the system behavior moves, albeit partially, from the user to the system. Thus, the potential for a confused user or unsatisfactory user experience increases.

A natural question that must now be addressed concerns the identification of criteria that an application might use as a basis for adapting its behavior. Context-aware computing (Schmidt, Beigl & Gellersen, 1999) provides one intuitive answer to this question. The notion of context first arose in the early 1990s as a result of pioneering experiments in mobile computing systems. Though an agreed definition of context has still not materialized, it concerns the idea that an application should factor in various aspects of the prevailing situation when offering a service. What these aspects might be is highly dependent on the application domain in question. However, commonly held aspects of context include knowledge of the end-user, for example through a user model; knowledge of the surrounding environment, for example through a geographic information system (GIS) model; and knowledge of the mobile device, for example through a suitably populated database. Other useful aspects of an end-user's context include an understanding of the nature of the task or activity currently being engaged in, knowledge of their spatial context, that is, location and orientation, and knowledge of the prevailing social situation. Such models can provide a sound basis for intelligently adapting system behavior. However, capturing the necessary aspects of the end-user's context and interpreting it is frequently a computationally intensive process, and one that may prove intractable in a mobile computing context. Indeed, articulating the various aspects of context and the interrelationships between them may prove impossible, even during system design (Greenberg, 2001). Thus, a design decision may need to be made as to whether it is worth working with partial or incomplete models of a user's context. And the benefit of using intelligent techniques to remedy deficiencies in context models needs to be considered in terms of computational resources required, necessary response time and the ultimate benefit to the end-user and service provider.

SOME REFLECTIONS ON CONTEXT

Mobile computing spans many application domains and within these, it is characterized by a heterogeneous landscape of application domains, individual users, mobile devices, environments and tasks (Figure 1). Thus, developing applications and services that incorporate a contextual component is frequently an inherently complex and potentially time-consuming endeavor, and the benefits that accrue from such an approach should be capable of being measured in some tangible way. Mobile computing applications tend to be quite domain specific and are hence targeted at specific end-users with specialized tasks or objectives in mind. This is in contrast to the one-size-fits-all attitude to general purpose software development that one would encounter in the broad consumer PC arena. For the purposes of this discussion, it is useful to reflect further on the following aspects of the average mobile user's context: end-user profile, devices characteristics, prevailing environment and social situation.

User Profile

Personalization and customization techniques assume the availability of sophisticated user models, and currently form an indispensable component of

Figure 1. An individual's current activity is a notoriously difficult aspect of an individual's context to ascertain with certainty

a number of well-known e-commerce related Web sites. Personalizing services for mobile computing users is an attractive proposition in many domains as it offers a promising mechanism for increasing the possibility that the end-users receive content that is of interest to them. Though this objective is likewise shared with owners of e-commerce sites, there are two issues that are of particular importance when considering the mobile user. Firstly, mobile interactions are almost invariably short and to the point. This obligates service providers to strive to filter, prioritize, and deliver content that is pertinent to the user's immediate requirements. The second issue concerns the question of costs. Mobile users have to pay for services, which may be charged on a KB basis, thus giving mobile users a strong incentive to curtail their use of the service in question if dissatisfied.

A wide number of features and characteristics can be incorporated into user models. As a basic requirement, some information concerning the user's personal profile, for example, age, sex, nationality and so on, is required. This basic model may then be augmented with additional sub-models that become increasingly domain-specific. In the case of standard e-commerce services, a record of the previous purchasing history may be maintained and used as a basis for recommending further products. Electronic tourist guides would require the availability of a cultural interest model, which as well as indicating cultural topics of interest to the user, would also provide some metric that facilitated the prioritization of their cultural interests.

Device Characteristics

Announcements of new devices are occurring with increasing frequency. Each generation successively increases the number of features offered, some of which would not be associated with traditional mobile computing devices, embedded cameras and MP3 players being cases in point. Though offering similar features and services, there are subtle differences between different generations, and indeed interim releases within the same generation, that make the life of a service provider and software professional exceedingly difficult and frequently irritating. From an interface perspective, screen size and support for various interaction modalities are two notable ways in which devices differ, and these have particular implications for the end-user experience. This problem is well documented in the literature and a number of proposals have been put forward to address this, the plasticity concept being a notable example (Thevenin & Coutaz, 1999). Other aspects in which mobile devices differ include processor, memory and operating system; all of which place practical limitations on what is computationally feasible on the device.

Prevailing Environment

The notion of environment is fundamental to mobile computing and it is the dynamic nature of prevailing environment in which the mobile user operates that most distinguishes mobile computing from the classic desktop usage paradigm. As an illustration, the case of the physical environment is now considered, though this in no way diminishes the importance of the prevailing electronic infrastructure. Scenarios in which mobile computing usage can occur are multiple and diverse. The same goes for physical environments. Such environments may be hostile in the sense that they do not lend themselves to easily accessing electronic infrastructure such as telecommunications networks. Other environments may experience extreme climatic conditions thus causing equipment to fail.

Developing a service that takes account of or adapts to the local physical environment is an attractive one. Two prerequisites are unavoidable, however. A model of the environment particular to the service domain in question must be available, and the location of the end-user must be attainable. In the former case, the service provider must construct this environmental model, possibly an expensive endeavor in terms of time and finance. In the latter case, an additional technological solution must be engaged—either one based on satellites, for example GPS, or one that harnesses the topology of the local wireless telecommunications networks. Each solution has its respective advantages and disadvantages, and a practical understanding of each is essential. However, by fulfilling these prerequisites, the service provider is in a position to offer services that take the end-users' physical position into account. Indeed, this vision, often termed location-aware computing (Patterson, Muntz & Pancake, 2003), has grasped the imagination of service providers and end-users alike. In essence, it is a practical example of just one single element of an end-user's context being interpreted and used as a basis for customizing services.

Social Situation

Developing a service that adapts to the end-user's prevailing social context is fraught with difficulty, yet is one that many people would find useful. What exactly defines social context is somewhat open to interpretation but in this case, it is considered to refer to the situation in which end-users find themselves relevant to other people. This is an inherently dynamic construct and capturing the prevailing social situation introduces an additional level of complexity not encountered in the contextual elements described previously.

In limited situations, it is possible to infer the prevailing social situation. Assuming that the end-user maintains an electronic calendar, the detection of certain keywords may hint at the prevailing social situation. Examples of such keywords might include lecture, meeting, theatre and so on. Thus, an application might reasonably deduce that the end-user would not welcome interruptions, and, for example, proceed to route incoming calls to voicemail and not alert the end-user to the availability of new email. Outside of this, one has to envisage the deployment of a suite of technologies to infer social context. For example, it may be that a device, equipped with a voice recognition system, may be trained to recognize the end-user's voice, and on recognizing it, infer that a social situation is prevailing. Even then, there may be a significant margin of error; and given the power limitations of the average mobile device, running a computationally intensive voice recognition system continuously may rapidly deplete battery resources.

ARTIFICAL INTELLIGENCE IN MOBILE COMPUTING

Artificial intelligence (AI) has been the subject of much research, and even more speculation, for almost half a century by now. Though failing to radically alter the world in the way that was envisaged, nevertheless, AI techniques have been successfully harnessed in a quite a number of select domains and their incorporation into everyday applications and services continues unobtrusively yet unrelentingly. Not surprising, there is significant interest amongst the academic community in the potential of AI for addressing the myriad of complexity that is encountered in the mobile computing area. From the previous discussion, some sources of this complexity can be easily identified. Resource management, ambiguity resolution, for example, in determining contextual state and resolving user intention in multimodal interfaces, and adaptation, are just some examples. Historically, research in AI has focuses on various issues related to these very topics. Thus, a significant body of research already exists in some of the very areas that can be harnessed to maximum benefit in mobile computing scenarios. A detailed description of these issues may be found elsewhere (Krüger & Malaka, 2004).

One pioneering effort at harnessing the use of intelligent techniques on devices of limited computational capacity is the Ambient intelligence (AmI) (Vasilakos & Pedrycz, 2006) initiative. AmI builds on the broad mobile computing vision as propounded by the ubiquitous computing vision. It is of particular relevance to this discussion as it is essentially concerned with usability and HCI issues. It was conceived in response to the realization that as mobile and embedded artefacts proliferate, demands for user attention would likewise increase, resulting in environments becoming inhabitable, or more likely, people just disabling the technologies in question. In the AmI concept, IUIs are envisaged as playing a key role in mediating between the embedded artefacts and surrounding users. However, AmI does not formally ratify the use of any particular AI technique. Choice of technique is at the discretion of the software designer whose selection will be influenced by a number of factors including the broad nature of the domain in question, the requirements of the user, the capability of the available technology and the implications for system performance and usability.

Having motivated the need for AI technologies in mobile contexts, practical issues pertaining to their deployment can now be examined.

STRATEGIES FOR HARNESSING AI TECHNIQUES IN MOBILE APPLICATIONS

It must be reiterated that AI techniques are computationally intensive. Thus, the practical issue of actually incorporating such techniques into mobile applications needs to be considered carefully. In particular, the implications for performance must be determined as this could easily have an adverse effect on usability. There are three broad approaches that can be adopted when incorporating AI into a mobile application and each is now considered.

Network-Based Approach

Practically all mobile devices are equipped with wireless modems allowing access to data services. In such circumstances, designers can adopt a kind of client/server architecture where the interface logic is hosted on the mobile devices and the core application logic deployed on a fixed server node. The advantage of such an approach is that the designer can adopt the most appropriate AI technologies for the application in question. However, the effect of network latency must be considered. If network latency is significant, the usability of the application will be adversely affected. Likewise, data rates supported by the network in question must be considered. Indeed, this situation is aggravated when it is considered that a number of networks implement a channel sharing system where the effective data rate at a given time is directly proportional to the number of subscribers currently sharing the channel. It is therefore impossible to guarantee an adequate quality of service (QoS) making the prediction of system performance difficult. Often, the worst case scenario must be assumed. This has particular implications where the AI application on the fixed server node needs either a significant amount of raw data or a stream of data to process.

One key disadvantage of placing the AI component on a fixed server node concerns the issue of cost. There is a surcharge for each KB of data transferred across the wireless network, and though additional revenue is always welcome, the very fact that the subscriber is paying will affect their perception of application in question and make them more demanding in their expectations.

A network–based AI approach is by far the most common and has been used in quite a number of applications. For example, neural networks have been used for profiling mobile users in conversational interfaces (Toney, Feinberg & Richmond, 2004). InCa (Kadous & Sammut, 2004) is a conversational agent that runs on a PDA but uses a fixed network infrastructure for speech recognition.

Distributed Approach

In this approach, the AI component of the service may be split between the mobile device and the fixed network node. The more computationally expensive elements of the service are hosted on the fixed network node while the less expensive elements may be deployed on the device. Performance is a key limitation of this approach as the computational capacity of the devices in question as well as the data-rates supported by the wireless network can all contribute to unsatisfactory performance. From a software engineering perspective, this approach is quite attractive as distributed AI (DAI) is a mature research discipline in its own right; and a practical implementation of DAI is the multi-agent system (MAS) paradigm.

One example of an application that uses a distributed approach is Gulliver's Genie (O'Grady & O'Hare, 2004). This is a tourist information guide for mobile tourists, realized as a suite of intelligent agents encompassing PDAs, wireless networks and fixed network servers. Agents on the mobile device are responsible for manipulating the user interface while a suite of agents on the fixed server collaborate to identify and recommend multimedia content that is appropriate to the tourist's context.

Embedded Approach

As devices grow in processing power, the possibility of embedding an AI based application on the actual physical device becomes ever more feasible. The key limitation is performance, which is a direct result of the available hardware. This effectively compromises the type of AI approach that can be usefully adopted. Overtime, it can be assumed that the capability and variety of AI techniques that can be deployed will increase as developments in mobile hardware continue and the demand for ever-more sophisticated applications increases. From an end-user viewpoint, a key advantage of the embedded approach concerns cost as the number of connections required is minimized.

One example of an application that uses the embedded approach is iDorm (Hagras et al., 2004), a prototype AmI environment. This environment actually demonstrates a variety of embedded agents including fixed motes, mobile robots and PDAs. These agents collaborate to learn and predict user behavior using fuzzy logic principles and, based on these models, the environment is adapted to the inhabitant's needs.

Deployment Considerations

Technically, all three approaches are viable, but the circumstances in which they may be adopted vary. For specialized applications, the networked AI approach is preferable as it offers greater flexibility and maximum performance, albeit at a cost. For general applications, the embedded approach is preferable, primarily due to cost limitations, but the techniques that can be adopted are limited. The distributed approach is essentially a compromise, incorporating the respective advantages and disadvantages of both the networked and embedded approach to various degrees. Ultimately, the nature of the application domain and the target user base will be the major determinants in what approach is adopted. However, in the longer term, it is the embedded approach that has the most potential as it eliminates the negative cumulative effect of network vagrancies, as well as hidden costs. Thus, for the remainder of this chapter, we focus on the embedded approach and consider how this might be achieved.

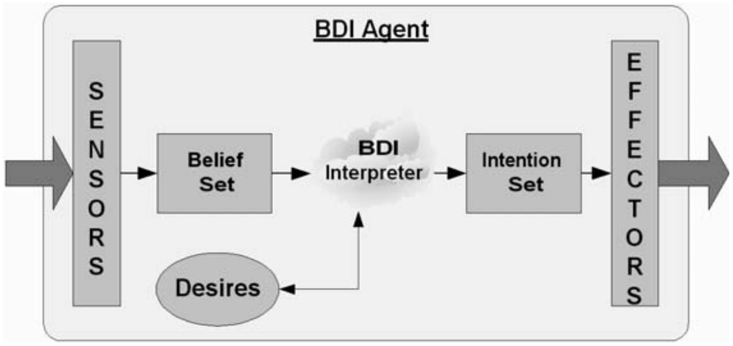
So what AI techniques can be adopted, given the inherent limitations of mobile devices? Various techniques have been demonstrated in laboratory conditions but one paradigm has been demonstrated to be computationally tractable on mobile devices: intelligent agents. As well as forming the basis of mobile intelligent information's systems, a number of toolkits have been made available under open source licensing conditions thus allowing software engineers access to mature platforms at minimum cost. Before briefly considering some of these options, it is useful to reflect on the intelligent agent paradigm.

THE INTELLIGENT AGENT PARADIGM

Research in intelligent agents has been ongoing since the 1970s. Unfortunately, the term agent has been interpreted in a number of ways thereby leading to some confusion over what the term actually means. More precisely, the characteristics that an arbitrary piece of software should possess before applying the term agent to it are debatable. In essence, an agent may be regarded as a computational entity that can act on behalf of an end-user, another agent or some other software artefact. Agents possess a number of attributes that distinguish them from other software entities. These include amongst others:

- **Autonomy:** The ability to act independently and without direct intervention from another entity, either human or software-related
- **Proactivity:** The ability to opportunistically initiate activities that further the objectives of the agent

Figure 2. Architecture of a BDI agent



- **Reactivity:** The ability to respond to events perceived in the agent's environment;
- **Mobility:** The ability to migrate to different nodes of a network as the need to fulfill its objectives dictates; and
- **Social ability:** The ability to communicate with other agents using a shared language and ontology leading to shared or collaborative efforts to achieve individual and shared objectives.

To what extent an agent possesses or utilizes each of those attributes is at the discretion of the designer. For clarity purposes, it is useful to consider agents as existing on a scale. At the lower end are so-called reactive agents. Such agents act in a stimulus-response manner, and a typical usage scenario might involve the agent monitoring for user interaction and reacting to it. Such agents are generally classified as weak agents (Wooldridge & Jennings, 1995). At the other end of the scale are so-called strong agents. Such agents maintain a sophisticated model of their environment, a list of goals or objectives, and plans detailing how to achieve these objectives. Such agents support rational reasoning in a collaborative context and are usually realized as multi-agent systems (MAS). This strong notion of agenthood is synonymous with the view maintained by the AI community.

One popular interpretation of the strong notion of agency is that of the belief-desire-intention (BDI) paradigm (Rao & Geogeff, 1995). This is an intuitive and computationally tractable interpretation of the strong agency stance. To summarize: beliefs represent what the agent knows about its environment. Note that the term environment can have diverse meanings here and may not just relate to the physical environment. Desires represent the objectives of the agent, and implicitly the raison d'être for the application. However, at any moment in time, an agent may be only capable of fulfilling some of its desires, if even that. These desires are then formulated as intentions and the agent proceeds to fulfill these intentions. The cycle of updating its model of the environment, identifying desires that can be fulfilled, and realizing these intentions is then repeated for the duration of the agent's lifecycle (Figure 2).

When should agents be considered for realizing a software solution? Opinion on this is varied. If the solution can be modeled as a series of dynamic interacting components, then agents may well offer a viable solution. However, many see agents as being particular useful in situations that are inherently complex and dynamic as their native capabilities equip them for handling the myriad of situations that may arise. Naturally, there are many situations that fulfill the criteria but, for the purposes of this discussion, it can be easily seen that the mobile computing domain offers significant opportunities for harvesting the characteristics of intelligent agents.

Intelligent Agents for Mobile Computing

As the capability of mobile devices grew, researchers in the intelligent agent community became aware of the feasibility of deploying agents on such devices, and perceived mobile computing as a potentially fertile area for the intelligent agent paradigm. A common approach was to extend the functionality of existing and well-documented MAS environments such that they could operate on mobile devices. It was not necessary to port the entire environment on to the device; it was just necessary to develop an optimized runtime engine for interpreting the agent logic. In this way, the MAS ethos is persevered and such an approach subscribes to the distributed AI approach alluded to previously. A further benefit was that existing agent-oriented software engineering (AOSE) methodologies could be used. In the case of testing, various toolkits have been released by the telecommunications manufacturers that facilitate the testing of mobile applications. A prudent approach is of course to test the application at various stages during its development on actual physical devices, as this will give a more accurate indication of performance, the look and feel (L&F) of the application and so on. For a perspective on deploying agents on mobile devices, the interested reader should consult Carabelea and Boissier (2003).

While a number of environments may be found in the literature for running agents on mobile devices, the following toolkits form a useful basis for initial consideration:

1. **LEAP (Lightweight Extensible Agent Platform)** (Bergenti, Poggi, Burg, et al., 2001) is an extension of the well-documented JADE platform (Bellifemine, Caire, Poggi et al., 2003). It is FIPA (http://www.fipa.org/) compliant and capable of operating on both mobile and fixed devices.
2. **MicroFIPA-OS** (Laukkanen, Tarkoma & Leinonen, 2001) is a minimized footprint of the FIPA-OS agent toolkit (Tarkoma & Laukkanen, 2002). The original FIPA-OS was designed for PCs and incorporated a number of features that did not scale down to mobile devices. Hence, MicroFIPA-OS minimizes object creation, reduces computational overhead and optimizes the use of threads and other resource pools.
3. **AFME (Agent Factory Micro Edition)** (Muldoon, O'Hare, Collier & O'Grady, 2006) is derived from Agent Factory (Collier, O'Hare, Lowen, & Rooney, 2003), a framework for the fabrication and deployment of agents that broadly conform to the BDI agent model. It has been specifically designed for operation on cellular phones and such categories of devices.
4. **JACK** is, in contrast to the three previous frameworks, a commercial product from the Agent Oriented Software Group (http://www.agent-software.com). It comes with a sophisticated development environment, and like AFME, conforms to the BDI agent model.

A detailed description of the each of these systems is beyond the scope of this discussion. However, the interested reader is referred to (O'Hare, O'Grady, Muldoon & Bradley, 2006) for a more advanced treatment of the toolkits and other associated issues.

FUTURE TRENDS

As mobile devices proliferate, and each generation surpasses its predecessor in terms of raw computational capacity and supported features, the potential for incorporating additional AI techniques will increase. In a similar vein, new niche and specialized markets for mobile services will appear. If a more holistic approach is taken towards mobile computing, it can be seen that developments in sensor technologies, fundamental to the ubiquitous and pervasive vision, will follow a similar trajectory. Indeed, the possibility of deploying intelligent agents on sensors is being actively investigated in widespread expectation that the next generation of sensors will incorporate processors of a similar capability to the current range of PDAs. Such a development is essential if the AmI vision to reach fruition.

As the possibility of incorporation of ever more sophisticated AI techniques increases, the potential for extending and refining the adaptability and IUI constructs for the support of mobile users increase. Indeed, adaptability may reach its fulfillment through the incorporation of autonomic computing precepts (Kephart & Chess, 2003). Self-configuring, self-healing, self-optimizing and self-protecting are the key attributes of an autonomic system, and it can be seen that incorporation of AI techniques may make the realization of these characteristics more attainable.

Finally, the practical issues of engineering mobile AI solutions must be considered. Mobile computing poses significant challenges to the traditional software engineering process, and the broad issue of how best to design for mobile services still needs to be resolved. The situation is exacerbated when AI technologies are included. However, it may be envisaged that as experience and knowledge of the mobile computing domain deepens and matures, new methodologies and best practice principles will emerge.

CONCLUSION

Mobile computing scenarios are diverse and numerous, and give rise to numerous challenges that must be overcome if the end-user experience is to be a satisfactory one. IUIs offers one viable approach that software designers can adopt in their efforts to make their systems more usable in what is frequently a hostile environment. However, the pragmatic issue of realizing mobile applications that incorporate intelligent techniques is of critical importance and gives rise to significant technical and design obstacles.

In this chapter, the broad issue of realizing an intelligent solution was examined in some detail. At present, the intelligent agent paradigm offers an increasingly viable proposition for those designers who wish to include intelligent techniques in their designs. To illustrate the issues involved, the intelligent agent paradigm was discussed in some detail.

As mobile developments continue unabated, the demand for increasingly sophisticated applications and services will likewise increase. Meeting this demand will pose new challenges for software and HCI professionals. A prudent and selective adoption of intelligent techniques may well offer a practical approach to the effective realization of a new generation of mobile services.

ACKNOWLEDGMENT

The authors gratefully acknowledge the support of the Science Foundation Ireland (SFI) under Grant No. 03/IN.3/1361.

REFERENCES

Bellifemine, F., Caire, G., Poggi, A., & Rimassa, G. (2003). *JADE—A white paper.* Retrieved January 2007 from http://jade.tilab.com/papers/2003/WhitePaperJADEEXP.pdf

Bergenti, F., Poggi, A., Burg, B., & Caire, G. (2001). Deploying FIPA-compliant systems on handheld devices. *IEEE Internet Computing, 5*(4), 20-25.

Collier, R.W., O'Hare, G.M.P., Lowen, T., & Rooney, C.F.B., (2003). Beyond prototyping in the factory of agents. In J. G. Carbonell & J. Siekmann (Eds.), *Lecture Notes In Computer Science*, 2691, 383-393, Berlin: Springer.

Carabelea, C., & Boissier O. (2003). Multi-agent platforms on smart devices: Dream or reality. Retrieved January, 2007, from http://www.emse.fr/~carabele/papers/carabelea.soc03.pdf.

Greenberg, S. (2001). Context as a dynamic construct. *Human-Computer Interaction*, 16, 257-268.

Hagras, H., Callaghan, V., Colley, M., Clarke, G., Pounds-Cornish, A., & Duman, H. (2004).

Creating an ambient-intelligence environment using embedded agents. *IEEE Intelligent Systems*, *19*(6) 12-20.

Höök, K. (2000). Steps to take before intelligent user interfaces become real. *Interacting with Computers*, *12*, 409-426.

Kadous, M.W., & Sammut, C. (2004). InCA: A mobile conversational agent. *Lecture Notes in Computer Science*, 3153, 644-653. Berlin: Springer.

Kephart, J.O., & Chess, D.M. (2003). The vision of autonomic computing. *IEEE Computer*, *36*(1), 41-50.

Kruger, A., & Malaka, R. (2004). Artificial intelligence goes mobile. *Applied Artificial Intelligence*, *18*, 469-476.

Langley, P. (1997). Machine learning for adaptive user interfaces. *Lecture Notes in Computer Science*, 1303, 53-62. Berlin:Springer.

Laukkanen, M., Tarkoma, S., & Leinonen, J. (2001). FIPA-OS agent platform for small-footprint devices. *Lecture Notes in Computer Science*, 2333, 447-460. Berlin:Springer.

Mitrovic, N., Royo, J. A., & Mena, E. (2005). Adaptive user interfaces based on mobile agents: Monitoring the behavior of users in a wireless environment. *Proceedings of the Symposium on Ubiquitous Computation and Ambient Intelligence* (pp. 371-378), Madrid: Thomson-Paraninfo.

Muldoon, C., O'Hare, G.M.P., Collier, R.W., & O'Grady, M.J. (2006). Agent factory micro edition: A framework for ambient applications, *Lecture Notes in Computer Science*, 3993, 727-734. Berlin:Springer.

O'Connor, A., & Wade, V., (2006). Informing context to support adaptive services, *Lecture Notes in Computer Science*, 4018, 366-369. Berlin: Springer.

O'Grady, M.J., & O'Hare, G.M.P. (2004). Just-in-time multimedia distribution in a mobile computing environment. *IEEE Multimedia*, *11*(4), 62-74.

O'Hare, G.M.P., O'Grady, M.J., Muldoon, C., & Bradley, J.F. (2006). Embedded agents: A paradigm for mobile services. *International Journal of Web and Grid Services*, *2*(4), 355-378.

Patterson, C.A., Muntz, R.R., & Pancake, C.M. (2003). Challenges in location-aware computing. *IEEE Pervasive Computing*, *2*(2), 80-89.

Rao, A.S., & Georgeff, M.P. (1995). BDI agents: from theory to practice. In V. Lesser and L. Gasser (Eds.), *Proceedings of the First International Conference on Multiagent Systems* (pp. 312-319). California: MIT Press.

Rhodes, B.J., Minar, N. & Weaver, J. (1999). Wearable computing meets ubiquitous computing: Reaping the best of both worlds. *Proceedings of the Third International Symposium on Wearable Computers* (pp. 141-149). California: IEEE Computer Society

Schmidt, A., Beigl, M., & Gellersen, H-W, (1999). There is more to context than location. *Computers and Graphics*, *23*(6), 893-901.

Tarkoma, S., & Laukkanen, M. (2002). Supporting software agents on small devices. In *Proceedings of the First International Joint Conference on Autonomous Agents and Multi-Agent Systems (AAMAS)* (pp. 565-566). NewYork: ACM Press.

Thevenin, D., & Coutaz, J. (1999). Plasticity of user interfaces: Framework and research agenda. In (M. A. Sasse & C. Johnson (Eds.), *Proceedings of IFIP TC 13 International Conference on Human-Computer Interaction (INTERACT'99)*, (pp. 110-117). Amsterdam: IOS Press.

Toney, D., Feinberg D., & Richmond, K. (2004). Acoustic features for profiling mobile users of conversational interfaces, *Lecture Notes in Computer Science*, 3160 (pp. 394-398). Berlin: Springer.

Vasilakos, A., & Pedrycz, W. (2006). *Ambient intelligence, wireless networking, ubiquitous Computing.* Norwood:Artec House.

Weiser, M. (1991). The computer for the twenty-first century. *Scientific American*, *265*(3), 94-100.

Wooldridge, M., & Jennings, N.R. (1995). Intelligent agents: Theory and practice. *Knowledge Engineering Review*, *10*(2), 115-152.

KEY TERMS

Ambient Intelligence: (AmI) was conceived by the Information Society Technologies Advisory Group (ISTAG) as a means of facilitating intuitive interaction between people and ubiquitous computing environments. A key enabler of the AmI concept is the intelligent user interface.

BDI Architecture: The Belief-Desire-Intention (BDI) architecture is an example of a sophisticated reasoning model based on mental constructs that can be used by intelligent agents. It is allows the modeling of agents behaviors in an intuitive manner that complements the human intellect.

Context: Context-aware computing considers various pertinent aspects of the end-user's situation when delivering a service. These aspects, or contextual elements, are determined during invocation of the service and may include user profile, for example language, age, and so on. Spatial contextual elements, namely location and orientation, may also be considered.

Intelligent Agent: Agents are software entities that encapsulate a number of attributes including autonomy, mobility, sociability, reactivity and proactivity amongst others. Agents may be reactive, deliberative or hybrid. Implicit in the agent construct is the requirement for a sophisticated reasoning ability, a classic example being agents modeled on the BDI architecture.

Intelligent User Interface: Harnesses various techniques from artificial intelligence to adapt and configure the interface to an application such that the end-user's experience is more satisfactory.

Mobile Computing: A computer usage paradigm where end-users access applications and services in diverse scenarios, while mobile. Mobile telephony is a popular realization of this paradigm, but wearable computing and telematic applications could also be considered as realistic interpretations of mobile computing.

Multi-Agent System: A suite of intelligent agents, seeking to solve some problem beyond their individual capabilities, come together to form a multi-agent system (MAS). These agents collaborate to fulfill individual and shared objectives.

Ubiquitous Computing: Conceived in the early 1990s, ubiquitous computing envisages a world of embedded devices, where computing artefacts are embedded in the physical environment and accessed in a transparent manner.

Chapter XXI
Tools for Rapidly Prototyping Mobile Interactions

Yang Li
University of Washington, USA

Scott Klemmer
Stanford University, USA

James A. Landay
University of Washington & Intel Research Seattle, USA

ABSTRACT

We introduce informal prototyping tools as an important way to speed up the early-stage design of mobile interactions, by lowering the barrier to entry for designers and by reducing the cost of testing. We use two tools, SUEDE and Topiary, as proofs of concept for informal prototyping tools of mobile interactions. These tools address the early stage design of two important forms of mobile interactions: speech-based and location-enhanced interactions. In particular, we highlight storyboarding and Wizard of Oz (WOz) testing, two commonly used techniques, and discuss how they can be applied to address different domains. We also illustrate using a case study: the iterative design of a location-enhanced application called Place Finder using Topiary. In this chapter we hope to give the reader a sense of what should be considered as well as possible solutions for informal prototyping tools for mobile interactions.

INTRODUCTION

The iterative process of prototyping and testing has become an efficient way for successful user interface design. It is especially crucial to explore a design space in the early design stages before implementing an application (Gould et al., 1985). Informal prototyping tools can speed up an early-stage, iterative design process (Bailey et al., 2001; Klemmer et al., 2000; Landay et al., 2001; Li et al., 2004; Lin et al., 2000). These tools are aimed at lowering the barrier to entry for interaction designers who do not have technical backgrounds, and automatically generating early-stage prototypes that can be tested with end users. The informal look and feel of these tools and their fluid input techniques, for example using pen sketching (Landay et al., 2001), encourage both designers and end users to focus on high level interaction ideas rather than on design or implementation details (e.g., visual layouts or colors). These details are often better addressed at a later stage. In this chapter, we focus on informal tool support for the early stage design of interactive mobile technologies. In particular, we describe informal prototyping tools that we developed for two increasingly important forms of mobile interaction: speech-based interactions (Klemmer et al., 2000) and location-enhanced interactions (Li et al., 2004).

The first of these two types of interactions, speech-based, works well on mobile phones, the major platform of mobile computing. These devices often have tiny screens and buttons to increase mobility, which makes speech interaction an important alternative. Although the accuracy of speech recognition is an important concern for a successful speech-based UI, the real bottleneck in speech interface design is the lack of basic knowledge about user "performance during computer-based spoken interaction" (Cohen et al., 1995). Many interaction designers who could contribute to this body of knowledge are excluded from speech design by the complexities of the core technologies, the formal representations used for specifying these technologies, and the lack of appropriate design tools to support iterative design (Klemmer et al., 2000). SUEDE (Klemmer et al., 2000) demonstrates how tool support can be used in the early stage design of speech-based user interfaces.

The second of these two types of interactions, location-enhanced, is important because of its implicit nature. While the explicit input channels (e.g., keyboarding or mouse pointing) available on mobile technology are more limited than on the desktop, the bandwidth of implicit input (using contextual information) is greatly expanded on mobile platforms. Mobile technology is more available in our context-rich, everyday lives than traditional desktop computing. One especially promising form of context-aware computing that has begun to see commercialization is location-enhanced computing, applications that leverage one's current location as well as the location of other people, places, and things (Li et al., 2004). For example, mobile phone services allow users to locate friends and family (LOC-AID), provide real-time navigation (InfoGation) and monitor and motivate users toward their fitness goals by using phone-based GPS to measure the user's speed, distance and elevation (BonesInMotion). E911 transmits a mobile phone user's current location when making emergency calls. However, location-enhanced applications are hard to prototype and evaluate. They employ sophisticated technologies such as location tracking and their target environment is mobile and in the field. Topiary (Li et al., 2004) demonstrates how high-level tool support can be provided for lowering the threshold and cost for designers to design and test location-enhanced applications.

Using SUEDE and Topiary as proofs of concept, we highlight two techniques commonly used in informal prototyping tools: storyboarding and Wizard of Oz (WOz) testing. To overcome the technical barrier for design, both SUEDE and Topiary employ a storyboarding-based approach for specifying interaction logic. To allow easy testing of prototypes, both tools employ WOz approaches where a human wizard simulates a sophisticated, nonexistent part of the prototype such as location tracking or speech recognition. To demonstrate how these types of tool can actually help prototype and test mobile technology, we introduce a case study using Topiary to design the Place Finder application.

BACKGROUND

User interface tools have been a central topic in HCI research. An extensive review of user interface tools can be found in (Myers et al., 2001). A large number of research prototypes and commercial products have been developed for rapid prototyping of user interfaces (Apple, 1987; Bailey et al., 2001; Hartmann et al., 2006; Klemmer et al., 2000; Landay et al., 2001; Li et al., 2004; Lin et al., 2000; Macromedia; MacIntyre et al., 2004).

In particular, informal prototyping tools are aimed at the early stages of a design process, and are used to create early-stage prototypes for testing key design ideas rather than building full-fledged final systems (Landay et al., 2001). They often result in example-based interface mockups that are able to demonstrate exploratory interactive behaviors but ignore other non-exploratory aspects of a desired system. Informal tools have shown great potential to facilitate the early stages of a design process and have been developed for various domains. For example, SILK is a tool for designing graphical user interfaces (Landay et al., 2001) that allows designers to create GUI prototypes by sketching and storyboarding. DENIM (Lin et al., 2000), a tool for the early stage design of Web sites, has become one of the most popular informal prototyping tools (downloaded over 100,000 times since 2000). Informal prototyping tools are often grounded in current practices of designers, e.g., paper prototyping (Rettig, 1994; Snyder, 2003), and lower the barrier to entry by maintaining the affordance of an existing practice. At the same time, informal tools provide extra value by allowing the easy editing and maintenance of a design, and by generating testable prototypes.

MAIN FOCUS OF THE CHAPTER

In our research, two features have emerged as being particularly valuable for rapidly prototyping mobile interactions. The first is storyboarding, which is inspired by traditional paper prototyping where designers draw key interaction flows visually on paper. Storyboarding is enhanced by electronic tool support to create the states and transitions. Many systems have been influenced by Harel's Statecharts model (Harel, 1987). Storyboarding is employed by both SUEDE and Topiary to lower the technical barrier for creating early-stage prototypes.

The second valuable feature is Wizard of Oz (WOz) testing, where a designer simulates part or all of the application logic by manipulating the interface in response to user input. This significantly reduces the time and labor required to create a testable prototype. As both speech-based interfaces and location-enhanced computing involve a necessary but sophisticated component, that is speech recognition and location tracking, respectively, both SUEDE and Topiary employed a WOz approach to avoid the complexity of introducing these components. To give an example of how this type of tool can help design and evaluate mobile technology in practice, we describe a case study for the iterative design of a PDA-based mobile Place Finder application using the Topiary.

Prototyping with Storyboards

In the early stages of design, it is important that tools allow designers to focus on the high-level concerns of interaction design, rather than forcing designers to also specify how these interactions are implemented. Storyboarding is an efficient way for designers to describe how a user interface should behave by enumerating concrete interaction sequences including both user input and interface output. These sequences should cover the key interaction paths of a proposed system in a particular design space. The concerns of early-stage prototyping are distinct from those of constructing an actual system, which focus more on completeness than exploration of the design space.

SUEDE allows two kinds of storyboarding: linear (conversation examples) and non-linear (design graphs of an actual interface) storyboarding. Designers start a design by creating simple conversation examples (see the Script Area at the top of Figure 1). These examples then evolve into the more complex, graph structure representing the actual interface design (see the design graph at the bottom of Figure 1) (Klemmer et al., 2000). The process of creating linear examples first and then forming more general design graphs is based on the existing practices of speech UI designers: we have found that often, designers begin the design process by writing linear dialog examples and then use those as a basis for creating a flowchart representation of the dialog flow on paper.

Designers lay out linear conversation examples horizontally as cards in the script area. *Prompts*, colored orange, represent the system's speech prompts. They are recorded by the designer for the phrases that the computer speaks. *Responses*, colored green, represent example responses of the end user. They are the phrases that participants make in response to prompts. System prompts alternate with user responses for accomplishing a task. A designer can record her own voice for the speech on both types of cards, as well as type in a corresponding label for each of the cards. By playing the recordings from left to right, the designer can both see and hear the example interaction. For example, in Figure 1, a designer has recorded a conversation example with the following alternating prompts and responses: "message from James," "erase it," "Are you sure," "Yes." After constructing example scripts, a designer can construct an actual design of a speech-based interface using the design graph (see Figure 1). A design graph represents a dialog flow based on the user's responses to the system's prompts. To create a design graph, designers can drag prompt or response cards from a script onto the design area, or create new cards on the design area, and link them into the dialog flow. SUEDE's storyboard mechanism embodies both the input and output of a speech interface in cards that can be directly manipulated (e.g., via drag & drop), and hides the complexity of using speech recognition and synthesis. This abstraction allows designers to focus on high-level design issues.

Topiary's storyboards also embed the specification of input and output interactions into a storyboard. Before introducing Topiary's storyboards, we first discuss Topiary's *Activity Map* workspace, a component designed for creating scenarios describing location contexts of people, places and things by demonstration (see Figures 2 and 3). The created scenarios can be used as input by Topiary storyboards when prototyping location-enhanced interactions (see Figure 4). Modeling implicit input, location context in this case, is a new challenge posed by mobile computing.

Topiary's Activity Map workspace employs an intuitive map metaphor for designers to demonstrate location contexts describing the spatial relationship of people, places and things. Designers can create graphical objects on the map to represent people, places and things (see Figure 2). Designers can move people and things on the map to demonstrate various spatial relationships. For example, in Figure 2, Bob is out of the library, the astronomy building and the café. However, Bob is close to the library because Bob's proximity region, indicated by the red circle around Bob, intersects with the library. The proximity region

Figure 1. SUEDE allows designers to create example scripts of speech-based interactions (top) and speech UI designs (bottom) by storyboarding.

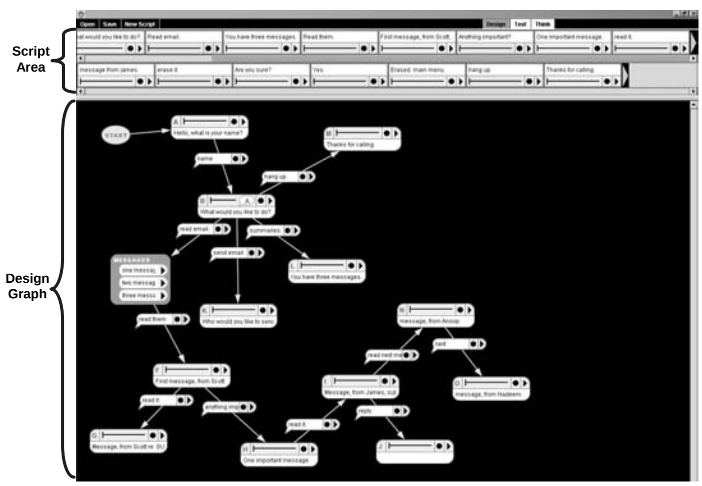


Figure 2. The active map workspace of Topiary is used to model location contexts of people, places and things and to demonstrate scenarios describing location contexts.

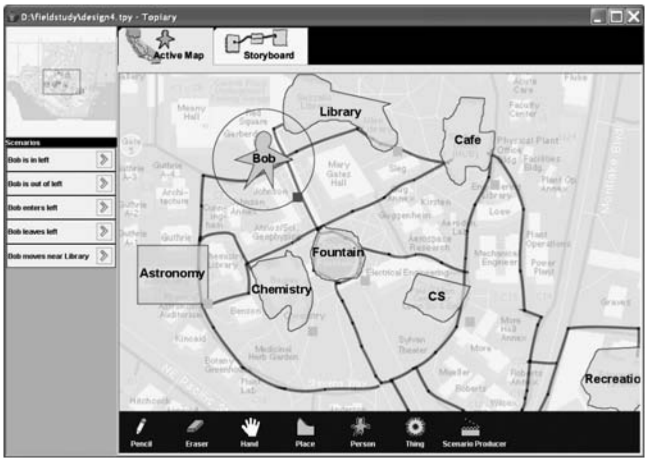
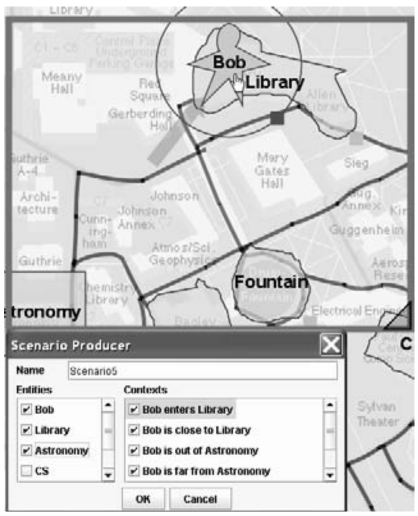


Figure 3. The designer drags Bob into the Library, with the context changing from "Bob is out of Library" to "Bob enters Library." As the entity CS (building) is unchecked, all related contexts to this place are filtered out.



can be resized by dragging the rectangular handle. These spatial relationships can be captured via Topiary's *Scenario Producer*. Like a screen capture tool, a designer can position a *Scenario Producer* window over entities of interest to capture spatial relationships (see Figure 3). A dialog box is then brought up that allows designers to select contexts of interest. Designers can demonstrate dynamic contextual transitions such as "entering" or "leaving" by moving entities within the recording window. For example, dragging Bob into the Library changes the event "Bob is out of Library" into "Bob enters Library" (see Figure 3).

Based on the location scenarios captured in the Active Map workspace, designers can create application prototypes in the Storyboard workspace (see Figure 4). In Topiary, a storyboard page represents a screen of visual output and a link represents transitions between pages. The key innovation in Topiary's storyboards is that scenarios created in the *Active Map* workspace can be used as conditions or triggers on links (Li et al., 2004). Designers create pages and links by sketching. Topiary has two kinds of links (see Figure 4). *Explicit links*, denoted in blue, start on ink within a page and they represent GUI elements that users have to click on, for example buttons or hyperlinks. *Implicit links*, denoted in green, start on an empty area in a page. They represent transitions that automatically take place when scenarios associated with that link occur. Explicit links model explicit interactions taken by end-users though they can be conditioned by sensed information, whereas implicit links model purely sensed data such as locations. One or more scenarios can be added to a link and multiple scenarios represent the logical AND of the scenarios. Multiple links starting from the same source represent the logical OR of transitions.

The Activity Map abstraction allows designers to focus on location contexts of interest rather than how these contexts can be sensed. Topiary's graphical storyboarding allows designers to specify rich interactions by drag & drop or sketching instead of specifying complex rules or Boolean logic expressions. From both SUEDE and Topiary, we conclude that the key to a successful informal tool is to devise an appropriate abstraction that matches designers' conceptual model for design and hides the less important aspects of exploring target interactions. Storyboards, as a meta-design metaphor, should be adapted and developed to fit within a specific domain when being applied.

Testing Using WOz Approaches

Speech-based or location-enhanced interactions resist rapid evaluation because the underlying technologies require high levels of technical expertise to understand and use, and a significant amount of effort to tune and integrate. For example, location-tracking infrastructures are not always available (e.g., GPS does not work well indoors) and they require a great deal of effort to deploy and configure. Incorporating these technologies too early in a design process may distract designers from fully exploring the design space. Consequently, we employed WOz approaches in these tools for testing early-stage prototypes. That is, a wizard (played by a designer or experimenter) simulates what these technologies would do in the final application.

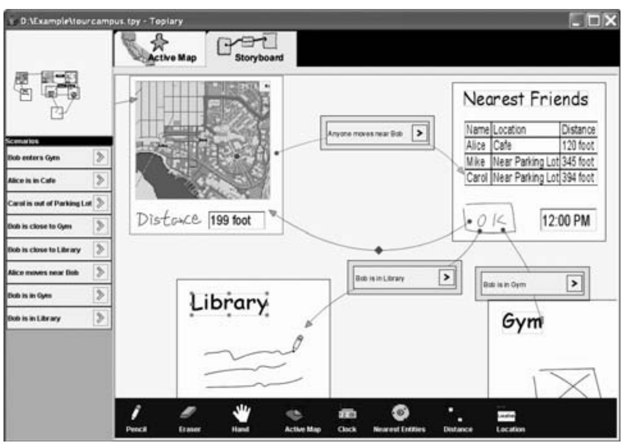


Figure 4. Topiary's Storyboard workspace allows application prototypes to be created. The lower three links (in blue) are explicit links, representing the behavior of the OK button depending where "Bob" is. The top link (in green) is an implicit link, representing an automatic transition from the Map page to the Nearest Friends page when "Anyone moves near Bob"

Wizard of Oz (WOz) testing has been widely employed in user interface design. In a WOz test, a wizard (often played by a designer) fakes an incomplete (or nonexistent) system component to conduct early user evaluations (Dahlbäck et al., 1993). In its most basic form, a WOz test works by the wizard simulating the machine behavior entirely. There is no computation in the loop at all. Examples of this form include testing paper prototypes by having the wizard physically move around the paper-based windows and menus (Rettig, 1994) and testing potential speech interface interaction flows by having a human operator on the other side of the telephone, following a pre-specified interaction graph. When an interactive prototype has been created (at least partially), the wizard can simply use the implemented interface. As a variant of this approach, a programmer can implement a functionally complete but suboptimal interface, and have the wizard control this interface during testing as a means of eliciting users' conceptual models of the task for example (Akers, 2006).

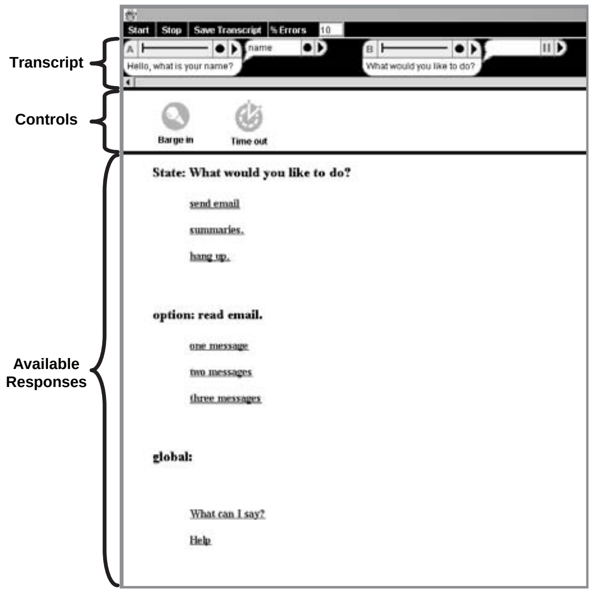
Significant gains beyond these basic approaches can be achieved through tools designed explicitly to support a Wizard of Oz approach. The fundamental insight behind a WOz-enabled tool is that the wizard is provided with a distinct user interface from that of the end user, and that the primary goal for this interface is to enable the wizard to rapidly specify to the system what the user's input was. In SUEDE, the interaction flow and audio prompts are specified by the designer ahead of time, and the user's responses to the speech prompts are interpreted by the wizard and specified to the system using a graphical interface that is runtime-generated based on the user's current state within the interaction flow. During a test, a wizard works in front of a computer screen. The participant performs the test away from the wizard, in a space with speakers to hear the system prompts and a microphone hooked up to the computer to record his responses. During the course of the test session, a transcript sequence is generated containing the original system audio output and a recording of the participant's spoken audio input.

When the wizard starts a test session, SUEDE automatically plays the pre-recorded audio from the current prompt. The wizard interface in SUEDE displays hyperlinks that represent the set of possible options for the current state (see Figure 5); the wizard waits for the test participant to respond, and then clicks on the appropriate hyperlink based on the response. Here, the wizard is acting as the speech recognition engine. Additionally, effective wizard interfaces should provide a display of the interaction history (as well as capture this for subsequent analysis); global controls for options generally available in an interface genre but independent of a particular interface or interface state (these globals can be defined by the tool or specified by the designer); and support for simulated recognition errors. This set of functionality enables the wizard to customize the test as she sees fit, handle user input beyond what was originally designed, and test whether the application is designed in such a way that users can understand and recover from "recognition errors."

Location-enhanced interfaces introduce the additional challenge that, almost by definition, a test must be conducted while moving to be ecologically valid. To address this, Topiary's WOz interface was specifically designed for a wizard to interact with the interface while walking. Topiary automatically generates user interfaces for testing, including the Wizard UI and the End-User UI, based on the Active Map and the Storyboard workspace. The Wizard UI (see Figure 6) is where a wizard simulates location contexts, as well as observes and analyzes a test. The End-User UI is what an end user interacts with during a test and it is also shown in the End User Screen window of the Wizard UI (see Figure 6) so that designers can monitor user interactions. The designer can also interact with the End-User Screen window for debugging purposes. The Wizard UI and End-User UI can be run on the same device (to let a designer try out a design) or on separate devices (one for the Wizard, the other for the user).

During a test, the wizard follows a user; each carries a mobile device, and these devices are connected over a wireless network. The wizard

Figure 5. SUEDE's Test mode is presented in a web browser and allows the wizard to focus on the current state of the UI (top) and the available responses for that state (bottom).



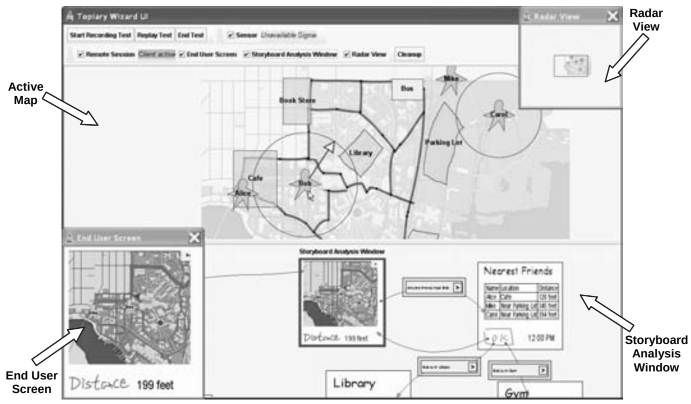
simulates location contexts by moving people and things around on the Active Map to dynamically update their location. The location changes of people and things on a map may trigger implicit transitions in the storyboard that will update the End-User UI. Topiary can also employ real location data if it is available, for more realistic testing at larger scales. A designer can choose to turn on a built-in location-tracking engine, based on Place Lab (LaMarca et al., 2005), which allows a WiFi-enabled or GSM-enabled device to passively listen for nearby access points to determine its location in a privacy-sensitive manner. In addition, a designer can analyze a design by recording a test and replaying it later. Topiary capture users' actions, like mouse movements and clicks, as well as physical paths traveled. The Storyboard Analysis window (see the bottom of Figure 6) highlights the current page and the last transition during a test or a replay session, which can help designers to figure out interaction flows.

Through our experience building SUEDE and Topiary, we have learned that effective tool support for Wizard of Oz testing comprises several key elements: the current state of the user interface (e.g., what is the current page in both tools), the current state of the user (e.g., the user's current location in Topiary) and the set of available actions (e.g., available responses in SUEDE). These elements should be provided to the wizard in an effective manner that allows the wizard to easily grasp and rapidly react. An effective Wizard interface should minimize the wizard's cognitive load by proactively maintaining a visible representation of state and having the displayed (and hence selectable) options for future action tailored to the state at hand.

A Case Study

To demonstrate how an informal prototyping tool can help at an early stage of the design process and how informal prototyping can inform the later design or development process, we report on our experience with the iterative design of a location-enhanced Place Finder using Topiary. A location-enhanced Place Finder embodies many features of location-enhanced, mobile applications. It allows users to find a place of interest more efficiently by leveraging the user's location (e.g., showing a

Figure 6. The Wizard UI has four major parts: The Active Map (a clone of the Active Map workspace) for simulating location contexts, the End User Screen for monitoring a user's interaction or debugging a design, the Storyboard Analysis Window for analyzing interaction logic and the Radar View for easy navigation of the Activity Map.



path to the destination). With the help of Topiary, we were able to efficiently explore the usability issues of map-based navigation techniques on a PDA held by a user walking in the field. Map-based navigation is a key component of a Place Finder application. Based on two design iterations that involved creating five different designs and testing them with four participants in the field as well as an analysis of implementation issues, we built a high fidelity prototype of the Place Finder.

The first iteration included four different user interface designs that shared the same the underlying map of places and paths in the Active Map workspace (see Figure 2). At each iteration, a user test was conducted in the field on a college campus, using a Toshiba Tablet PC and an HP iPAQ™ Pocket PC. During each test, the wireless communication between the two devices was based on a peer-to-peer connection so that the connection was not affected by the availability of access points in the field.

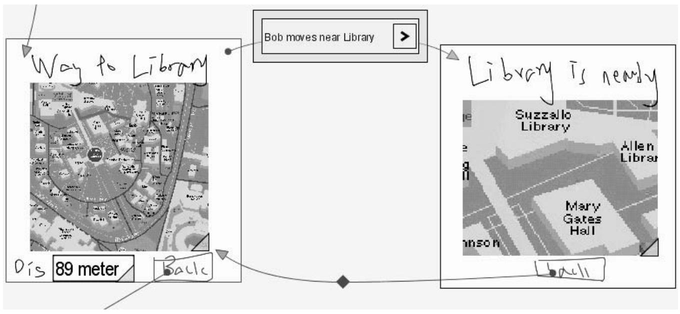
Iteration #1

It took us only three hours in total to create four prototypes, each using a different navigation technique. The first design shows a map of the entire campus (see Figure 7a). The second design shows an area centered on the user and lets the user manually zoom in and out (see Figure 7b). The third design uses the user's current location to show different regions of the campus (see Figure 7c). The last design is similar to the second, except it automatically zooms in or out based on the user's current speed (see Figure 7d). This last design was based on the idea that people are reluctant to interact with a device while walking. All four designs showed the user's current location and shortest path (see the thick pink lines in Figure 7) to the target, both of which are updated dynamically by Topiary.

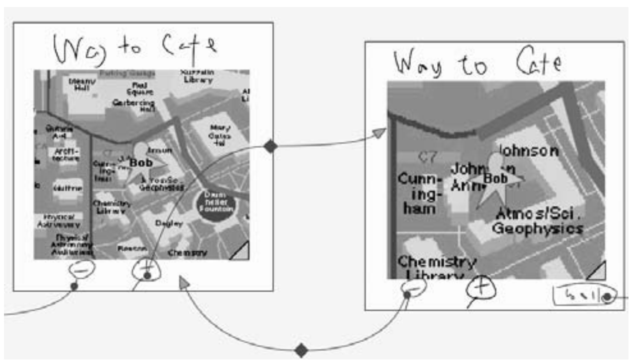
Four navigation segments were included in the test of Iteration #1, one segment for each of the four designs. These four segments were selected based on two principles. First, to smoothly connect the four experimental segments, the target of a segment should be the starting point of the following segment. Second, each segment should cover an area that requires a moderate walk, not too long or too short (e.g., an eight minute walk), and can produce a path with enough complexity to avoid simple paths (e.g., the entire path is a straight line.)

We had three participants try all four designs on a PDA in the field, with a wizard updating their location on a Tablet PC. Each experimental ses-

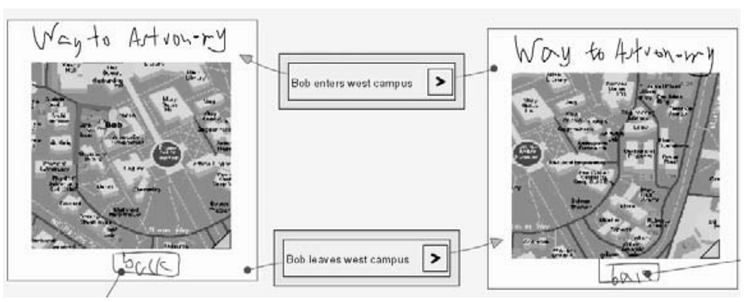
Figure 7. Storyboard fragments of the four designs in Iteration #1. A page, which holds maps and sketches, represents a screen of visual output of the user interface. Arrows (links) between pages represent transitions. The blue links represent GUI elements such as buttons for which scenarios can be used as conditions (not shown here). The green links represent transitions that can automatically take place when the associated scenarios occur.



(a) Design #1 shows the entire campus and a detailed map is automatically shown when a user gets close to a target. Here the scenario "Bob moves near library" triggers showing a detailed map around the library.

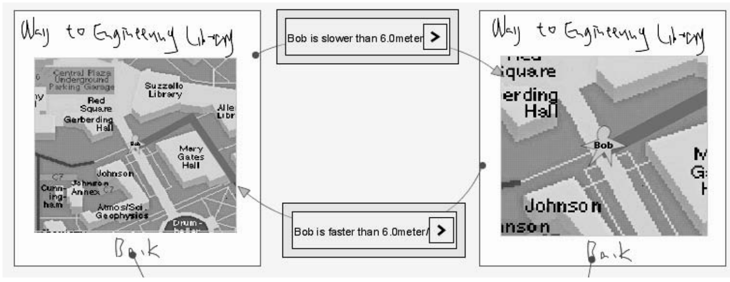


(b) Design #2 shows an area automatically centered on the user and lets the user manually zoom in or out by clicking on the sketched "+" or "–" buttons.



(c) Design #3 uses the user's current location to show different regions of the campus. Here the scenario "Bob enters (or leaves) west campus" triggers showing the west (or east) region of the campus.

continued on following page



(d) Design #4 is similar to Design #2 except it automatically zooms in or out based on the user's current speed. Here the scenario "Bob is slower (or faster) than 0.6 meter/s" triggers showing maps at different zoom levels.

sion lasted about one hour and each segment took about fifteen minutes to complete. During the test, we were able to make some minor changes to the design instantly in response to the participant's suggestions.

All three participants preferred the map centered on the user's current location (#2 and #4). The problem with the first design is that it shows the entire campus on a small PDA screen, which turned out to be hard to read. The third design does show more detail but it does not give a global view of the campus and the participants complained that they could not see the target until they were physically in that region, although they were still able to see the path.

Two participants preferred manual zooming to automatic zooming as they thought manual zooming gave them more control over the zoom levels. However, the other participant thought both kinds of zooming were good to use. All our participants thought the distance label from Design #1 was useful and they also suggested that we should flash the target when users get close to it.

One common problem with the four designs was that there was not enough orientation information provided. We originally thought users could figure out their orientation by referring to nearby buildings and the continuous change of their location on the map.

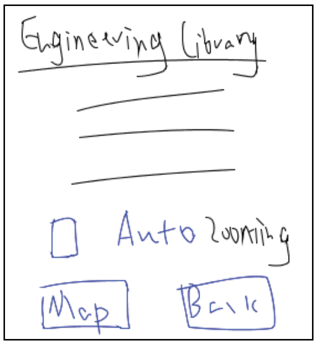
Iteration #2

Based on participant feedback and our observations during Iteration #1, we spent *one hour* creating a new design combining the best features of the four designs (see Figure 8). We added a page for users to choose automatic or manual zooming (see Figure 8a). We explored different ways of showing orientation on a map, including rotating the map, showing an orientation arrow, and showing trajectory arrows (see Figure 8b). These orientation representations are provided by Topiary. In addition, in response to the participants' request, we added the feature of flashing a target when it is nearby. We tested this new design again with three people[1]. Each test session lasted about half an hour in total. In the middle of the test, we turned on the sensor input that is built into Topiary to see how sensor accuracy affected our participants.

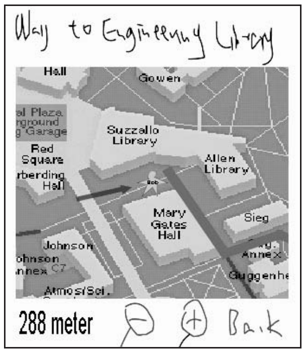
Our participants gave us many useful comments. For example, two of them suggested showing a movement trail to help to indicate orientation. Also, the inaccurate update of the user's location, either by the Wizard or by the sensor input (while it was turned on), did confuse the participants. As a result, one person suggested showing a region for the possible location instead of just a point. They also gave us some other suggestions, such as placing the distance label at the top of the screen rather than at the bottom.

Interestingly, some of our participants did not realize their location was being updated by a wizard rather than by real sensors. It was also observed that the prototype showed an optimal path to a participant who had spent three years on the campus but did not know the existence of this path. We did not know this path either and we simply drew a road network in Topiary by

Figure 8. Two screens (pages) of the new design



(a) A user can select or deselect the checkbox to choose automatic or manual zooming.



(b) A map screen with zooming buttons and a trajectory arrow

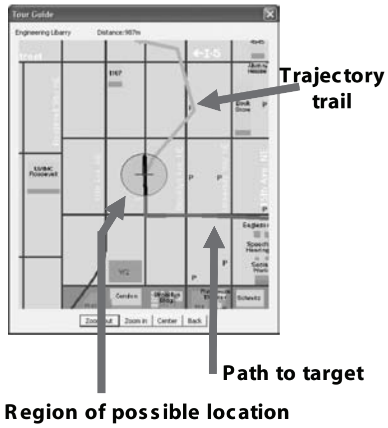
which this path was automatically constructed by the tool.

Building a High Fidelity Prototype

Through these two iterations of informal prototyping and testing, we got a rough view of what the Place Finder should be like. Then it was the time to consider implementation issues and to create a high fidelity prototype. Because we did not want to add an extra device, like GPS, for the Place Finder PDA, we chose to use Place Lab for location sensing, since it requires only WiFi. However, Place Lab, like GPS, cannot provide precise orientation. As a result, we decided to show a movement trail (feedback from the earlier study) instead of showing potentially inaccurate directional arrows or employing map rotation. In addition, because the movement speed cannot be accurately measured, we cut the automatic zooming feature, although one participant showed interest in it. This also helped improve application performance on the PDA.

Based on the earlier tests and an analysis of the implementation issues, we built a high fidelity prototype in Java, using the IBM J9 SWT Java toolkit, in about *two weeks* (see Figure 9). We have used this prototype in the field for hours and it has helped us to find places that we had never been to before. We also got positive feedback from people to whom we demoed this prototype. However, performance on the PDA is still a major issue with this prototype and more profiling is necessary before widely testing it.

Figure 9. The high fidelity prototype was built based on the informal prototypes and an analysis of implementation issues



Lessons Learned

This study offers lessons in two areas. First, it identified usability issues as well as solutions for building map-based navigation techniques in a location-enhanced Place Finder application. Second, the study gives an example of how early stage design iteration can be conducted using informal tools. The study showed obvious advantages over traditional paper prototyping since we were able to test our ideas in the field and leverage those results for later stage development.

Informal prototyping and testing in *hours* was much less expensive than directly building a high fidelity prototype over a period of *weeks* and then testing it with users. The tools allowed us to focus on interaction rather than implementation details. It turned out that little feedback from our participants was related to the informal look of the interface. *Focusing on key interactions* rather than specifying the behaviors of the entire application is important to efficiently conducting early stage design because prototyping tools often employ example-based approaches. In our study, only five places were modeled for testing the five low-fi designs. Once the early usability issues were solved, the design was scaled up to 35 places in the high fidelity prototype.

Carefully testing in the field is important for a successful early stage design because the field is where a mobile application design will be used. Testing in the field requires extra consideration when compared to controlled experiments in a lab setting. The *Wizard of Oz* technique was extremely useful in testing an early stage design since it can reasonably approximate realistic situations. On the other hand, using real sensor input, if not expensive, might help find more usability problems due to the inaccuracy of sensors in a test.

FUTURE TRENDS

Sensors such as accelerometers are becoming available on an increasing number of mobile devices to detect a user's context (e.g., movement, lighting or ambient noise) as well as other peripheral input (e.g., digital compass for the orientation of the device). With these sensors, richer interactions can be constructed. It is important for informal prototyping tools to support interaction design based on the available sensors of the platform. Multimodal interaction that combines multiple interaction modalities has shown promise. Speech-based interaction enhanced by location context is an extremely promising research direction. By leveraging location context, a system can optimize speech recognition by focusing on phrases that have meaning in a particular context. This brings new research opportunities to the rapid prototyping of mobile technology. The two tools that we discussed in this chapter address speech-based interaction and location-enhanced computing separately. It would be interesting to combine the strengths of these types of tools for prototyping location-enhanced speech user interfaces.

CONCLUSION

Informal prototyping tools play an important role in the early stage design of interactive mobile technology. They lower the threshold for entry and reduce the cost for prototyping and testing. As a proof of concept of informal prototyping tools for mobile interaction, we discussed how SUEDE and Topiary address the design of speech-based interaction and location-enhanced interaction, respectively, the two representative types of interaction for mobile technology. We highlight two common features of these tools: graphical storyboarding and Wizard of Oz testing. To show how these tools can help an iterative design process, we reported on our experience in iteratively prototyping a location-enhanced Place Finder application, and testing its prototypes with real users in the field. The study indicated that this type of tool allowed a designer to effectively explore a design space in the early stages of design. As mobile computing becomes more powerful and prevalent, there will be more opportunities for research on informal prototyping tools for the design and evaluation of interactive mobile technology.

REFERENCES

Akers, D. (2006). CINCH: A cooperatively designed marking interface for 3D pathway selection. In *UIST'06* (pp. 33-42).

Apple (1987). *HyperCard User's Guide*. Apple Computer, Inc.

Bailey, B. P., Konstan, J. A., & Carlis, J. V. (2001). DEMAIS: designing multimedia applications with interactive storyboards. In *ACM Multimedia* (pp. 241-250).

BonesInMotion. BiM Active. Retrieved from http://bonesinmotion.com/corp/

Cohen, P. R., & Oviatt, S. L. (1995). The role of voice input for human-machine communication. In *Proceedings of the National Academy of Sciences*, *92*(22), 9921-9927.

Dahlbäck, N., Jönsson, A. & Ahrenberg, L. (1993). Wizard of Oz studies—Why and how. In *Intelligent User Interfaces '93* (pp. 193-200).

Gould, J. D., & Lewis, C. (1985). Designing for usability: Key principles and what designers think. *Communications of the ACM*, *28*(3), 300-311.

Harel, D. (1987). Statecharts: A visual formalism for complex systems. *Science of Computer Programming*, *8*(3), 231-274.

Hartmann, B., Klemmer, S. R., Bernstein, M., Abdulla, L., Burr, B., Robinson-Mosher, A., & Gee, J. (2006). Reflective physical prototyping through integrated design, test, and analysis. In *UIST'06* (pp. 299-308).

InfoGation. Odyssey Mobile. from http://www.infogation.com/

Klemmer, S. R., Sinha, A. K., Chen, J., Landay, J. A., Aboobaker, N., & Wang, A. (2000). SUEDE: A wizard of oz prototyping tool for speech user interfaces. In *CHI Letters: 2*(2), *UIST'00* (pp. 1-10).

LaMarca, A., Chawathe, Y., Consolvo, S., Hightower, J., Smith, I. E., Scott, J., Sohn, T., Howard, J., Hughes, J. Potter, F., Tabert, J., Powledge, P, Borriello, G., & Schilit, B. N. (2005). Place lab: Device positioning using radio beacons in the wild. In *Proceedings of Pervasive'05* (pp. 116-133).

Landay, J. A., & Myers, B. A. (2001). Sketching interfaces: Toward more human interface design. *IEEE Computer, 34*(3), 56-64.

Li, Y., Hong, J. I., & Landay, J. A. (2004). Topiary: A tool for prototyping location-enhanced applications. In *CHI Letters: 6*(2), *UIST'04* (pp. 217-226).

Lin, J., Newman, M. W., Hong, J.I., & Landay, J. A. (2000). DENIM: Finding a tighter fit between tools and practice for Web site design. In *CHI Letters: 2*(1), *CHI'00* (pp. 510-517).

LOC-AID. LOC-AID People Service. from http://www.loc-aid.net/people_en.htm

MacIntyre, B., Gandy, M., Dow, S., & Bolter, J. D. (2004). DART: A Toolkit for Rapid Design Exploration of Augmented Reality Experiences. In *CHI Letters: 6*(2), *UIST'04* (pp. 197-206).

Macromedia. Director. from http://www.macromedia.com/software/director/

Myers, B., Hudson, S. E. & Pausch, R. (2001). Past, present and future of user interface software tools. In J. M. Carroll (Ed.), *The new millennium*, (pp. 213-233) New York: ACM Press, Addison-Wesley.

Rettig, M. (1994). Prototyping for tiny fingers. *Communications of the ACM*, *37*(4), 21-27.

Snyder, C. (2003). *Paper prototyping: The fast and easy way to design and refine user interfaces*. Morgan Kaufmann.

KEY TERMS

Graphical Storyboarding: A technique that informal prototyping tools often employ for designers to describe how an interface should behave. Like a state transition diagram (STD), it has the concepts of states and transitions. However, in graphical storyboarding these states and transitions represent high level UI components or events rather than the computational elements found in a traditional STD.

Informal Prototyping: A type of user interface prototyping used in the early stages of design in which designers explore a design space by focusing on key interaction ideas rather than visual (e.g., color or alignment) or implementation details. These details are often better considered when creating **hi-fidelity** prototypes at a later stage. Paper prototyping is a representative form of informal prototyping in which designers draw interfaces as well as interaction flows on paper.

Informal UI Prototyping Tools: A type of UI prototyping tool that fluidly supports an informal UI prototyping practice. These tools maintain an "informal" look and feel, use fluid input techniques (e.g., sketching) and can automatically generate testable, interactive prototypes.

Location-Enhanced Applications: Computer applications that leverage the location of people, places and things to provide useful services to users. For example, based on the user's current location, show the nearby restaurants or friends. By using the location context, this type of application reduces explicit input required from a user (such as mouse clicks or typing).

Sketch-Based User Interfaces: A type of user interface in which users interact with a computer system by drawing with a pen. The drawings can be recognized and interpreted as commands, parameters or raw digital ink. This type of interface has shown promise in supporting domains such as UI design, mechanical design, architectural design and note-taking.

Speech-Based Interfaces: A type of user interface in which the user input is submitted mainly via speech. A computer system responds based on either recognized words or vocal variation of the speech. The interface output is typically auditory (e.g., when it is on a phone) or visual.

User Interface Prototyping: A practice of creating user interface mockups to test some aspects of a target interactive system.

UI Prototyping Tools: Electronic tools supporting a user interface prototyping process.

Wizard of Oz Testing: A technique for testing an incomplete interface mockup, named after the movie the *Wizard of Oz*. In this technique, a wizard (often played by a designer) fakes an incomplete (or nonexistent) system component to conduct early user evaluations, (e.g., a wizard can simulate speech recognition when testing a speech-based interface or location tracking when testing a location-enhanced application).

Chapter XXII
Modelling and Simulation of Mobile Mixed Systems

Emmanuel Dubois
University of Toulouse III, France

Wafaa Abou Moussa
University of Toulouse III, France

Cédric Bach
University of Toulouse III, France

Nelly de Bonnefoy
University of Toulouse III, France

ABSTRACT

Interactive systems are no longer expected to be used in confined and predefined places. By increasingly taking advantage of the physical environment, interactive systems are becoming mixed, that is, merging physical and digital worlds. Moreover, they support user's mobility and thus can be referred to as "mobile mixed systems." To overcome technology-driven development processes and to take into account their physical nature and mobile dimensions, specific design approaches are required. From this perspective, we present the interweaving of an existing design model (ASUR) for mixed systems, and a 3-D environment (SIMBA) for simulating modelled mobile mixed system. The aims are to support the investigation of mobile mixed system design through the dedicated modelling approach, and to better understand the limit of the modelled solutions through their simulation. This constitutes a first step toward an iterative method of design for mobile mixed systems, based on "midfidelity" prototyping.

INTRODUCTION

Interacting with a computer system through keyboard, mouse, screen, and speaker in a fixed and predefined working environment is no longer the only available solution. Although such static interactive situations are useful for individual and acontextual applications, mobile interactive systems are tightly interwoven with the existing user's activity, physical artefacts, and application domain resources (Rodden, 1998). They provide a good support to information dissemination, opportunistic share or collection of information, intuitive manipulation, and so forth. Simple examples include the Wii console interaction device, museum guide overlaying exhibits with digital information, and so forth.

Recent advances in the technological, software, and communication infrastructure domains facilitate the implementation of different forms of mobile interactive systems (Renevier, 2004):

- Nomadic systems are carried by a mobile user. Wireless, light, and small devices are required.
- Ubiquitous systems offer services in any places, thus supporting the user's mobility, but remains invisible. It relies on the understanding of the use and manipulation of physical artefacts.
- Context-sensitive systems are influenced by physical properties of the user. User's location and orientation are the two main characteristics of importance.

In this chapter we particularly focus on the two last categories, and use the term "mobile mixed systems" to refer to them. The MARA prototype constitutes a good example of such systems (Kähäri & Murphy, 2006). In this denomination, the "mobile" dimension covers the third form listed (context-sensitive) and "mixed systems" depicts interactive systems that combine the use of physical and digital entities (Dubois, Nigay, Troccaz, Chavanon, & Carrat, 1999) and covers the second case (ubiquitous). Mixed systems include interaction paradigms such as tangible interfaces, augmented and mixed reality, and so forth.

The design of mobile mixed systems introduces many new aspects such as physical artefacts and properties description, links between physical and digital entities, identification of the multiple interaction facets, and so forth. However, traditional HCI approaches leave out most of these specific aspects. Therefore, most contributions in this domain consist of the development of empirical and ad hoc systems. Other contributions have recently introduced models and approaches supporting early design phases: physical objects and interaction description, context models, spatial organisation, and so forth.

The work presented in this chapter aims to take advantage of both kinds of approaches, empirical developments and modelling approaches, to support the design of mobile mixed systems. By coupling these approaches, the goal is twofold:

- To add some rationale in the currently ad hoc solutions in order to justify and document design choices, and to better support the reusability of technical implementation parts;
- To settle direct links between early design and development phases, in order to anchor the reasoning about interactive technique design and its software implementation.

To do so, we have chosen to adopt an iterative HCI design process: participatory design. As presented in Mackay et al. (Mackay, Ratzer, & Janecek, 2000), this iterative process is based on four steps, each of them instrumented with different methods:

- **Analysing:** It refers to the requirements analysis and problems identification. This step relies on user's observations *in situ* with probes, or *in vitro* within labs.
- **Designing:** It consists of generating ideas to solve aspects of the interaction techniques being designed. This second step can be based on a combination of informal techniques (brainstorming, focus group, interviews, etc.) and formal techniques (models, notation, diagrams, etc.) , respectively used to generate ideas and then describe the solutions.

- **Prototyping:** It leads to the development of a part of the interactive systems, covering one or more requirements or problems identified in the first step. Tools for prototyping include paper, video, or mock-ups, but also high-fidelity prototypes based on programming languages.
- **Evaluating:** It aims at studying usability aspects of the prototype, and potentially identifying new requirements or problems to trigger a new loop through the participatory design process. This last step involves user's test or speak aloud for example.

The first step is similar to traditional HCI observation and analysis. The second one must address specificities of a mobile mixed system and in particular, its physical nature. The third step must be adapted to this specific area: on one hand, usual low-fidelity prototyping tools (pen and papers, storyboard, etc.) do not reflect all the specificities of these systems (physical properties of the manipulated objects, the interaction context, user's position in the interactive environment …); on the other hand, high-fidelity prototyping tools are very costly (expensive equipment, complexity of software combination and communication, etc.). Finally, the fourth step is still in a very prospective state.

We therefore concentrate on the coupling of the two intermediary steps (design and prototyping), and propose to interweave 1) an extension of an existing modelling approach specific to mixed systems (ASUR) and, 2) a 3-D environment (SIMBA) used to simulate a mobile mixed system that has been previously modelled with ASUR.

This combination constitutes a support to the design and cheap, yet relevant, prototyping of a mobile mixed system. As a result it allows a designer of such interactive systems to:

- Investigate possible interaction techniques in a mobile mixed interaction context thanks to a systematic design model
- Better understand the limits of the modelled solution thanks to a cheap "midfidelity" prototype.

After a brief overview of existing mixed systems design approaches and simulation supports, we present the two facets of our contribution. We then sum up the main aspects of our mixed systems modelling approach, ASUR, and present its extension to suit the mobile mixed system specificities. In the following section, we introduce our ASUR-based simulation environment, called SIMBA, and illustrate the use of SIMBA on a concrete aeronautical application. Finally, we identify a set of perspectives and future trends for this work.

BACKGROUND

Designing Mixed Systems

Definition and Classification

In order to face the profusion of terms, such as augmented or mixed reality, tangible user interfaces, user's augmentation, and so forth, we introduced the generic term mixed systems. It includes all kinds of interactive systems involving physical and digital entities (Dubois et al., 1999). Different aspects have been used to compare mixed interaction techniques: type of data provided (Azuma, 1997; Feiner, MacIntyre, & Seligmann, 1993; Noma, Miyasato, & Kishino, 1996) (3-D graphics, sound, haptic), entity being enriched (Mackay et al., 1996) (user, objects, environment), information representation (Milgram & Kishino, 1994).

As opposed to these implementation-driven approaches, we refined mixed systems domain into two classes (Dubois et al., 1999) according to the type of interaction:

- **Augmented reality systems** (AR) enhance the interaction between the user and her/his physical environment by providing additional digital capabilities and/or information
- **Augmented virtuality systems** (AV) enhance the interaction between a user and a computer by enabling and interpreting the use of physical objects.

So far, no clear consensus has been raised to adopt a common and unique definition and similarly, a large number of prototypes have been developed to illustrate the technical feasibility to use and combine new technologies. To overcome this exploratory process, different design approaches have been investigated. They can be organised into two separate types: implementation support that aims at assisting the development of mixed systems, and modelling approaches that aim at improving the understanding and the exploration of mixed system design solution.

Implementation Support

Many ready-to-use libraries have been developed to integrate specific features of mixed systems such as video-based marker tracking (Kato & Billinghurst, 1999), gestures recognition (Hong & Landay, 2000) physical data sensing, and so forth.

More than a support to the integration of various technologies, development environments have been worked out. For instance, AMIRE (Haller, Zauner, Hartmann, & Luckeneder, 2003) and DWARF (Bauer, Bruegge, Klinker, MacWilliams, Reicher, Riß, Sandor, & Wagner, 2001) offer a set of predefined components, patterns, and connection facilities. Extension mechanisms are not clearly stated, but such approaches provide the developers with a structured view of the application.

Finally, additional works intend to connect these advances with existing standardised tools (Director), or format (SVG).

Modelling Approaches

TAC paradigm (Shaer, Leland, Calvillo-Gamez, & Jacob, 2004) and MCPrd (Ishii& Ullmer, 2000) architecture describe the elements required in tangible user interfaces: one focuses on the description of physical elements while the second focuses on the software structure of TUI.

Other models support the exploration of mixed systems design space: they are based on the identification of artefacts, entities, characteristics, and tools relevant to a mixed system. They are based on a set of HCI and context of use models (Trevisan, Vanderdonckt, & Macq, 2005), on the description of the different interaction facets of the user's interaction with a mixed system (Dubois, Gray, & Nigay, 2003), on the definition of the software elements required by the mixed modalities (Coutrix & Nigay, 2006).

More recent works in mixed systems try to link design and implementation steps by projecting scenarios on software architecture models (Delotte, David, & Chalon, 2004; Renevier et al., 2004) or combining Petri Nets and DWARF components (Hilliges, Sandor, Klinker, 2005).

High level of abstraction, component-based approach, tools interoperability, and implementation support constitutes the main challenges of today's mixed system design approaches. However, models remain static and do not easily address evolutions in the interactive situation, especially when the user is mobile. 3-D simulation is one cheap alternative to the final system development and gives a simple and realistic representation of the mobile dimension of mobile mixed systems.

3-D Simulation Environment

Among numerous fast 3-D generation platforms, we quote the popular VR Juggler (Cruz-Neira, Bierbaum, Hartling, Just, & Meinert, 2002), Avengo (Tramberend, 1999), VIPER (Caubet & Torguet, 1995), NPSNET, DIVE, Bamboo, and Blue-c (Rodriguez, Jessel, & Torguet, 2001). They can render distributed environments and enable interaction through various traditional input devices. However, generated 3-D worlds can rarely be considered as reusable.

In contrast, Balcisoy et al.'s (Balcisoy, Kallmann, Fua, & Thalmann, 2000) platform's main drawbacks are that multiple prototypes must be developed to test the usability of different interaction techniques, and most of the questions related to the interaction mode (voice commands, motion capture, etc.) remain untackled. With SPI (Bernard, Chevalier, & Baudoin, 2004) human behaviour is modelled, but other user's interaction dimensions are lacking. Finally, all these software tools cannot easily fit into a development process, thus making design and prototyping appear as distinct and separate tasks.

More recently, a modelling of an interactive situation can be used to simulate an interactive system displaying information on different surfaces (screen, wall, etc.) (Molina Masso, Vanderdonckt, Gonzalez Lopez, Caballero, & Lozano Perez, 2006). The user can freely modify the use and organisation of the surfaces. However, there is no way to study other interaction aspects than these surfaces position and no clear possibilities for populating the simulation environment with other components.

Using 3-D environments as a support for simulating systems constitutes one of the current challenges of interactive system designers. So far, no simulators are sufficient to support a complete simulation of a mobile mixed interaction system and offer extension mechanisms.

Before presenting and illustrating our simulation environment of mobile mixed systems, we present the UI design model on which the simulation platform is based.

ASUR: A MIXED SYSTEM MODELLING APPROACH AND AN EXTENSION

The ASUR model (Dubois et al., 2003) adopts a user's interaction point of view on the design of mixed systems. The following sections detail and illustrate this model, highlight some limits, and present an extension based on ergonomic criteria.

ASUR Basic Principles

To introduce the ASUR model, we consider an augmented museum scenario. Information is provided according to the exhibit in front of which the visitor is standing. To display this information, the visitor carries a PDA. Information is stored in a database to which the user is automatically connected when entering the museum. Finally, the museum is equipped with a localisation system.

The first step of the ASUR modelling consists in identifying entities involved in this task.

ASUR Components

ASUR distinguishes different component types:

- The **S component** depicts the computer system, including computational and storage capabilities (the museum database), and data acquisition and delivery.
- The **U component** refers to the user of the system: the visitor.
- $\mathbf{R_{object}}$ **and** $\mathbf{R_{tool}}$ **components** denote physical entities involved when performing the task. $\mathbf{R_{object}}$ designate real focus of the task such as the exhibit, and, $\mathbf{R_{tool}}$ play the role of intermediary entities required to perform the task, for example a wand manipulated by the user to point an exhibit of interest.
- $\mathbf{A_{in}}$ **and** $\mathbf{A_{out}}$ **components** represent adaptors conveying data from the physical to the digital world ($\mathbf{A_{in}}$, e.g. a localiser) or conversely ($\mathbf{A_{out}}$, e.g. a PDA).

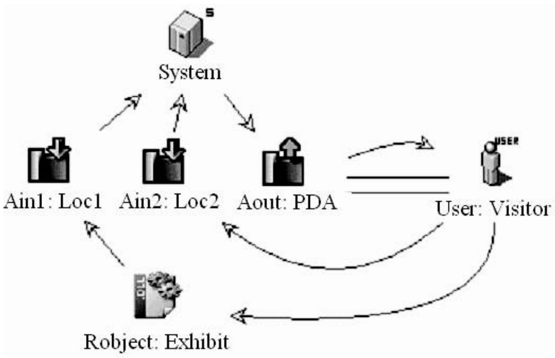
These components are not autonomous and need to communicate during the task realisation. Such communication is modelled with ASUR relationships.

ASUR Relationships

We identified three different types of ASUR relationships. A **data exchange (A→B)** means that component B may perceive information rendered by component A. In our example, the visitor observes the exhibit (R_{Object}→U) and data displayed on the PDA (A_{Out}→U). Two distinct devices localise the exhibit and the user (R_{Object}→A_{In1}, U→A_{In2}), and transmit the positions to the computer system (A_{In1}→S, A_{In2}→S). After processing the data, updates are sent to the PDA (S→A_{Out})

Physical proximity (A==B) denotes the physical link that exists between two entities. For example, it represents the fact that the user is holding the PDA in his hand.

Figure 1. ASUR modelling of the augmented museum



Triggers (A⇨B) is always linked to a data exchange (C→D): the data transfer from C to D will only occur when a specific spatial condition is reached between A and B. No such link is used in the museum example.

ASUR Characteristics

Figure 1 is a diagrammatic representation of this first part of the ASUR modelling of the augmented museum. Additional characteristics are used to refine this modelling (Dubois et al., 2003): **location** and **perception sense** indicate where the user has to focus to get the information and through which human sense it is perceivable, **dimension** (1-D, 2-D, 3-D) and **point of view** refine the description of information transfer.

ASUR Outcomes

ASUR constitutes a good support for comparing and reasoning about the specificities of mixed systems in early design phases. However, the S component includes every digital entity. Typically, it is not possible to accurately model ubiquitous interfaces: links between the physical nature of the system and its digital complement are hidden in relationships to a unique digital entity (S component), and every digital interaction facet takes its origin in the same global digital entity.

An ASUR refinement is thus required to develop a more precise model of ubiquitous systems that support user's mobile activity. We present this extension in the next section.

ASUR Extension

The main goal of this extension is to refine the S component that aggregates every digital aspect involved in a user's task with a mixed system. Refining this component requires the identification of the relevant objects involved in this component. According to our knowledge, no domain model describing the digital part of a mixed interactive system has been studied so far. Therefore, we chose to rely on a domain model of virtual environments. We present this domain model, the principles of its combination with ASUR, and the resulting ASUR extension.

Combining ASUR and a Virtual Environment Domain Model

Our definition of a virtual environment domain model has been triggered by the elaboration of a method for the definition and execution of user's experiments in the field of human virtual environment interaction (HVEI). To overcome the current lack of methodological support when performing experiments in HVEI and analysing the results, we adapted the existing ergonomic criteria (EC) (Bastien & Scapin, 1993) to the specificities of virtual environment (VE) (Bach & Scapin, 2003). The adaptation process is similar to the one already used and validated in the context of Web applications (Leulier, Bastien, & Scapin, 1998). The first phase of this adaptation consisted in a thorough analysis of the literature that led to the compila-

Figure 2. The eight main groups of the list of elements constituting a virtual environment (V.E.)

1 – User's profile: user's preferences and abilities management
2 – Represented objects: general description and role of the objects present in the EV
3 – Actions: availability of services on objects
4 – Spatial organisation of the V.E.
5 – 3D Decor of the V.E.
6 – 3D Boundaries of the V.E.
7 – Autonomous elements
8 – V.E. behaviour: general description and actions of the V.E.

tion of 170 ergonomic recommendations, ordered according to a set of 20 ergonomic criteria. The second phase consisted in identifying elements to which these recommendations applied. Finally, we organised the elements identified through this process to produce a list of elements constituting a virtual environment: the list is structured into eight main groups listed in Figure 2.

Since at least one ergonomic recommendation is attached to each of these 73 elements, elements of the list are relevant when performing an ergonomic analysis of an interactive situation with a virtual environment: a model of interactive systems involving a virtual environment must then include these elements. Mixed systems being partly physical, partly digital (or virtual), integrating this list of elements into our ASUR model will improve its ability to model the digital part, that is, the S component.

To proceed with this extension, we first identified among the 73 elements those which corresponded to existing component or characteristics of the ASUR model: "Actions" (3rd point) refers to the ASUR relationships, "Represented objects" (2nd point) includes the interaction mode (gesture, haptic, audio, etc.) and corresponds to the characteristic "perception sense" of the ASUR adapters, and so forth.

We then focused on the remaining elements of the list to refine our model by integrating these complementary elements.

ASUR Model Refinement

The analysis of the list of elements constituting a V.E. resulted into two main changes in the initial ASUR model: the first one is a split in the S component that leads to the identification of three distinct components and their own characteristics; the second one consists in introducing a new kind of relationship.

Split in the S Component

Initially this component included every digital aspect involved when interacting with the system. The first split consists in separating the fundamental digital resources such as the operating system, drivers, storage capacity, rendering and communication capabilities, and so forth, from the digital entities relevant for the interaction with the system.

Former aspects are now grouped into the "***S meta-componen,***" which is implicitly included in every ASUR model. Indeed these aspects do not have an impact on the user's interaction with the system or cannot be directly modified by the designer. Describing them in more detail is thus postponed to the software design step.

To better describe the latter aspects, we first rely on a traditional approach in HCI that consists in separating the interaction into an execution and a perception phase (Norman, 1986). As a result, we introduce two different kinds of S components: S_{tool} and $S_{presentation}$ components.

S_{tool} component. S_{tool} depicts a digital entity used as a tool to perform an action. Its activation has an effect on another digital entity. It corresponds to an articulatory subtask required to apply the desired action onto a digital object. For example, a menu offering several commands to apply to a 3-D object is considered as an S_{tool}: selecting the command is not the finality of the interaction, it is rather intended to apply the command on the 3-D object.

The list of elements constituting a V.E. highlighted two relevant characteristics for the S_{tool} components. They were both derived from the second items of the list ("Represented objects"), and its precise nature (simple, group, etc.) and role (input interaction or navigation support).

$S_{presentation}$ component. $S_{presentation}$ symbolise digital entities carrying information relevant to the user's interaction. The set of $S_{presentation}$ components present in an ASUR model forms the digital-object domain: it represents the concepts manipulated by the application kernel. The analysis of the list of elements constituting a V.E. highlighted subtype of $S_{presentation}$ components:

- **S_{object} component:** Symmetrical to the R_{object} component, it designates a digital object of the task, the object with which it is associated, and the purpose of the task. For example, in a word-processing situation, the file containing the document constitutes the digital object of the task.
- **S_{info} components:** They represent domain objects relevant to the task being performed, without being the focus of the task. Such components contain information related to the internal state of some part of the system: when perceivable, they help the user to build a correct mental representation of the state of the system. In the word-processing example, the amount of pages of the document is an S_{info} component.

The list of elements constituting a V.E. highlighted two relevant characteristics for the S_{info} and S_{object} components. Subitem of "represented objects" reveals the need for the S_{object} and S_{info} to precise if they accept commands or information from other digital component or not. The second important characteristic is derived from the second, fourth, fifth, sixth, and eighth items, and is again the role played by a S_{info} in the interactive task: it can be one of interaction feedback, help, decors/boundaries, or data.

An Additional Relationship

ASUR relationships are used to represent physical relations (proximity or trigger) or data exchanges between two components. Relationships between physical and digital entities are mediated by an adaptor and correspond to input and output of the interactive system. For example in the case of the mediaBlocks (Ullmer & Ishii, 1999), a user (U) manipulates physical cubes (R_{tool}) to modify the sequence order of video sequences (S_{object}): a localiser (Ain) is used to detect the position of the cubes and apply the changes to the video sequences ($R_{tool} \rightarrow A_{in}, A_{in} \rightarrow S_{object}$). This part of the modelling describes concisely data flows between entities involved in the interaction, but does not highlight that the digital video sequences are physically represented by the physical cubes. This aspect is a design-significant aspect of a user's interaction with a mixed system, and has to be added to the initial model.

The "**Representation link**" is thus a new kind of ASUR relationship that connects a physical and a digital entity. It does not represent any information exchange, but depicts a semantic link between two entities. Characteristics of this relationship include the dynamic or static aspect of its existence and the analogy, in term of representation and behaviour of the entity, and the one representing it. These characteristics have been identified on the basis of the list of elements constituting a V.E., the "reproduction fidelity" property introduced by Milgram and Kishino (1994) and the link described by Renevier (2004).

This extended version of the ASUR model, presented in a more concise way with the ASUR Metamodel (Dupuy-Chessa & Dubois, 2005), now supports the detailed modelling of the user's interaction when performing a task with an interactive system merging physical and digital artefacts, such as mobile and ubiquitous systems. Based on this model and its relationship to ergonomic criteria, ergonomic properties studies may also be performed (Dubois et al., 2003). However, testing some physical aspects of the interaction, such as the environment constraints, the availability of the entities, and so forth, is still not easy to conduce on a static model. We thus develop SIMBA, an environment that supports the simulation in a 3-D space of a mobile mixed system modelled with ASUR.

SIMBA: AN ASUR-BASED 3-D SIMULATION ENVIRONMENT

Our primary goal is to develop a prototyping platform that enables closer observation of some aspects of the user's interaction with a mobile mixed system through simulation. To build upon past experiences in terms of simulation and overcome their limitations, we identified a set of requirements:

- Taking advantage of the outcomes of earlier design phases to seamlessly integrate this prototyping approach into the development cycle.
- Avoiding the need for an extensive knowledge in 3-D programming in order to let the designer concentrate on usability rather than technical issue solving.
- Including extension mechanisms for inserting new simulation elements.

In the following sections we present SIMBA functionalities, the software model on which SIMBA's components are based and an illustration of its use on a concrete application.

SIMBA Overall Process

Loading the Required Information

SIMBA stands for "**SIM**ulation **B**ased on **A**sur." Its use is based on:

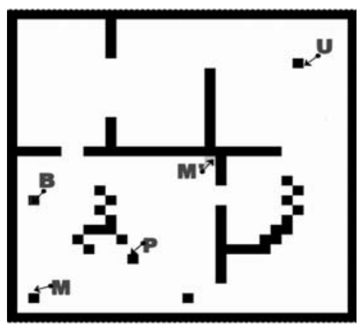
- An ASUR model and a representation of the physical organisation of the interactive environment
- Predefined SIMBA elements that simulate physical objects, devices, or digital objects during the simulation

GUIDE-Me, the graphical environment for the manipulation of ASUR models (Guide-Me, 2004), generates XML descriptions of ASUR models. Such a file is loaded in SIMBA and interacting entities, as well as relationships between them, are extracted. One unique ID is attached to each entity.

Secondly, a 2-D map is loaded: it represents the physical environment in which the interaction takes place. The position of any pixel encodes the position of one entity in this environment. The RGB-coded colour of the pixel is used to express additional information: the Red value holds the ID provided by the XML description of the ASUR model, the Green and Blue values identify the existing SIMBA element to associate

Figure 3. Extract of an XML file corresponding to an ASUR model (left) and the associated 2-D map (right)

```
- <component>
    <name>Localisateur</name>
    <id>3</id>
    <description />
  - <diagramPosition>
      <xPosition>109.0</xPosition>
      <yPosition>132.0</yPosition>
    </diagramPosition>
  - <charactComponent>
      <shareLevel>1</shareLevel>
      <location>Salle</location>
    </charactComponent>
  - <typeOfComponent>
    - <adapter>
      - <charactAdapter>
        - <sensAction>
            <physicalAction />
            <language>Physical action</language>
          </sensAction>
          <sensPerception>visuel</sensPerception>
        </charactAdapter>
      - <adapterType>
        - <adapterIn>
            <charactAdapterIn />
          </adapterIn>
        </adapterType>
      </adapter>
    </typeOfComponent>
  </component>
```



to this ASUR entity during the simulation. Finally, SIMBA interface dialogs allow the user to provide additional information.

Starting and Using the Simulation

Once the simulation is started, SIMBA generates a Doom-like 3-D environment. The designer using the SIMBA simulation environment manipulates an avatar, inside this virtual world, through the standard input devices. SIMBA also handles collision detection. Simulated visual output devices, such as monitors, projectors, PDAs, and so forth, are represented in a separate 2-D window, the "output emulation window," to enable WIMP-style interaction with menus and dialogs. Interaction between the 3-D environment and the output emulation window is handled by SIMBA.

Figure 4 illustrates the simulation of a visitor in the augmented museum modelled in ASUR in Figure 1.

Running a simulation thus entirely relies on the loading of an ASUR model of the system to simulate and a 2-D map representing the physical settings, including the positioning of the different objects involved. In order to facilitate the development of additional SIMBA elements, a SIMBA element model has been adopted and is presented in the next section.

SIMBA Element Model

The implementation of SIMBA, our 3-D simulation environment, is based on Java and Java3D as a rendering engine. This provides the portability and the high level of programming essential to simplify the development of additional SIMBA entities usable in the simulation.

Following most conceptual decompositions in the domain (Sanchez-Segura, 2005), SIMBA elements include two components:

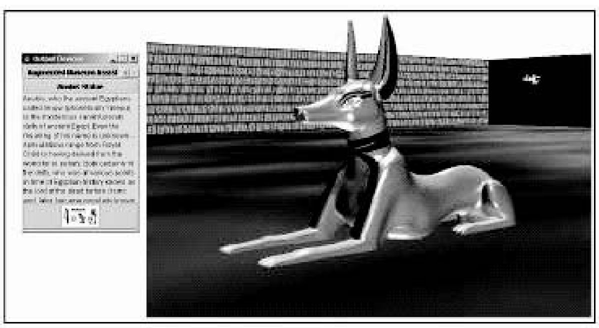
- **"Presentation Components":** Usually programmed by 3-D experts, they handle the loading and rendering of the objects.
- **"Functionality Components":** Usually programmed by interaction experts and application domain specialists, they are in charge of the data processing and communication between the different entities.

Next sections detail the different facets of the functionality component.

Interaction Between SIMBA Entities

To handle the communication between entities, SIMBA uses the relationships provided by the ASUR model. Let us consider a camera (A_{in}) that localises a user ($U \rightarrow A_{in}$) and transmits this information to a digital component ($A_{in} \rightarrow S_{object}$) in charge of storing the user's path.

Figure 4. SIMBA simulation output: visitor's view on museum (right) and "output emulation window" (left)



The functionality component associated with this camera is developed to respond to entity movements inside its visual range, and to identify the moving entity. If the detected entity corresponds to the user, the functionality component associated with the camera posts an event that is caught by SIMBA and forwarded to each entity connected to the camera via an ASUR relationship (the S_{object} in our case). The callback mechanism is triggered inside the receiving entity's functionality component, and the HandleAsurPost() method is automatically invoked upon receipt of the event. This method only receives a reference to the event emitter.

As a result, developing a new functionality component only requires writing the HandleAsurPost() method, that is, the code to execute when this component receives information from another ASUR component. The core of SIMBA environment is in charge of managing the broadcasting of events with respect to the relationships expressed in the ASUR model.

Additional Functionalities

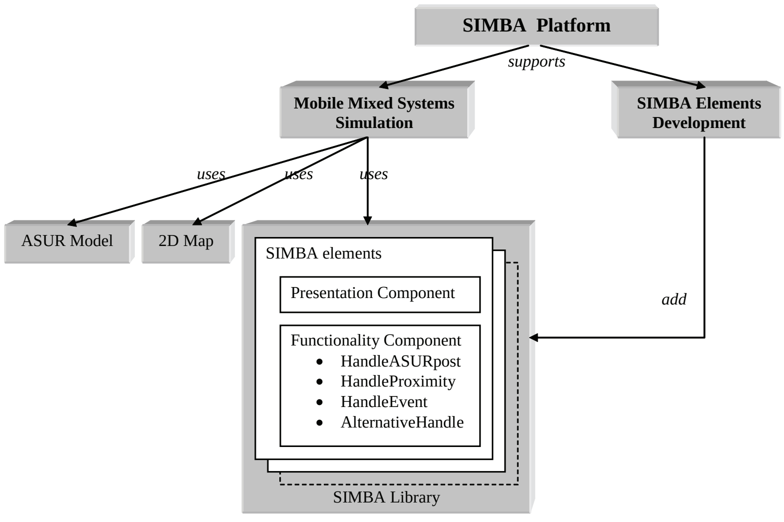
Other methods are included in the functionality component model to handle other kinds of events. Three methods may be automatically invoked by SIMBA, based on the triggering event:

- HandleProximity()is required to simulate physical world constraints that are not clearly expressed by the ASUR model. For example, when the avatar comes close enough as to satisfy certain physical constraints (i.e., to press a button, it must be within the user's arm reach), this method will have to activate or deactivate the device.
- HandleEvent() is called to respond to JAVA events independent of the user's interaction inside the simulated environment.
- AlternativeHandle() is the default method invoked whenever a condition different from those described above is met (e.g., Java advanced functionality).

Most functionality components will only implement one of the mentioned methods. Using distinct function calls for different types of events leads to a simpler event handling

Any entity of a mobile mixed system implemented with SIMBA adheres to this model. The SIMBA element library is thus made of a hierarchical set of SIMBA elements. Of course presentation and functionality component of each element are subclassable, thus promoting code-reuse.

Figure 5. Different aspects of the SIMBA platform



A Concrete Illustration

Aeronautical Maintenance Context

The case study takes place in the aeronautical maintenance process. Maintenance activities are expanding into many aspects of the aeronautical domain. The maintenance process is thus getting richer and more complex; as a result, the cognitive workload of the operators is extremely high. To reduce this workload and provide a better support to the maintenance process, ubiquitous and mixed information systems are promising because they are able to provide mobile operators with appropriate information, at the right place and time and, in the best suitable way.

Such systems are based on physical artefacts (element of planes, tools, manuals, etc.) and the sensing of contextual parameters. Among these dynamic aspects, the localisation of the operator is required to provide information relevant to the maintenance process, such as availability of services (wifi, localisation, etc.) and collaborators. However, before implementing such a system, design decisions must be taken with respect to:

- The technology (type and physical organisation)
- The rendered information (content and representation)
- Data flows among entities

To validate such design decisions, full-scale experiments are highly suitable but remain very expensive: many human and material resources in the operating maintenance areas are required, technologies are not easy to implement, planes immobilisations are too expensive, and so forth.

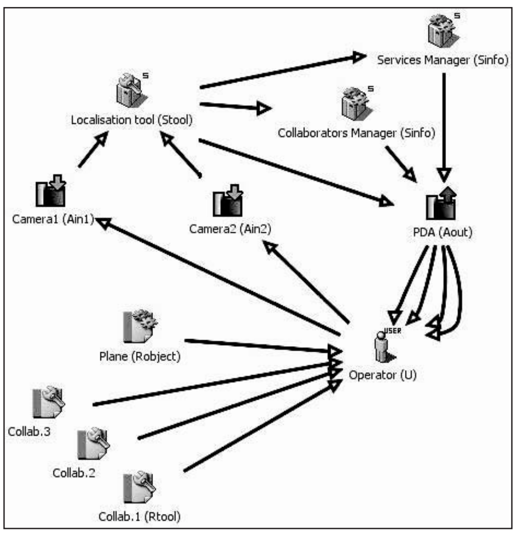
To overcome the high cost of such experiments, we used SIMBA to study the first two design aspects mentioned: impact of the use of cameras rather than RF-ID to localise the operator, and the nature of the representation used to visualise available services. Next section illustrates the whole simulation process.

ASUR Modelling of the Aeronautical Maintenance Situation

This mobile mixed system involved different ASUR components listed and represented in Figure 6:

- The operator in charge of the maintenance (U)
- The plane on which the maintenance is performed (R_{object})

Figure 6. ASUR model of an aeronautical maintenance situation involving a mobile operator



- Collaborators, that is, persons helping the operator during the maintenance process (set of R_{tool}), but not localised in this simulation
- Two cameras detecting the operator's position and transmitting it to the system ($A_{in}1$, $A_{in}2$)
- A PDA displaying to the operator, appropriate computer-generated information (A_{out})
- A localisation tool that converts localisation data coming from the cameras into coordinates useful for the computer system (S_{tool})
- A services manager in charge of identifying the available services ($S_{info}1$)
- A collaborators manager in charge of identifying the available collaborators ($S_{info}2$)

Simulation with SIMBA

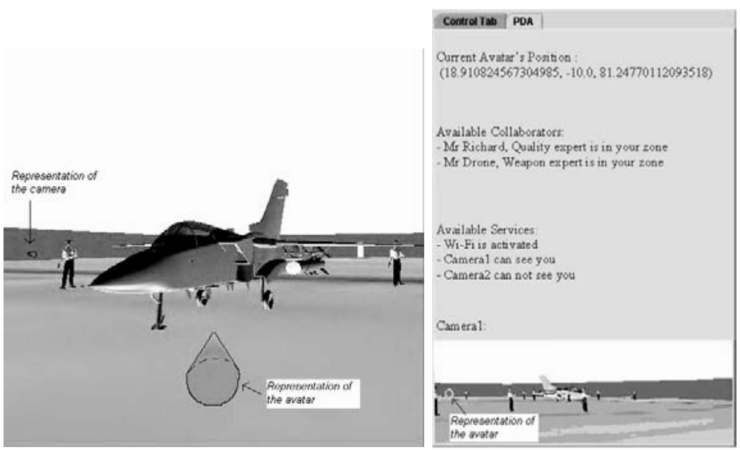
To prototype this situation into our 3-D simulation environment, three SIMBA elements were specifically developed: the localisation tool and the services and collaborators managers (S_{tool}, $S_{info}1$, $S_{info}2$). The SIMBA element corresponding to the PDA derives from an existing class of PDA to which a video display capability has been added. To simulate the cameras, two instances of an existing SIMBA_Camera element are used. Finally, 3DS max models are attached to the presentation components associated to the planes and the collaborators.

Once the simulation is started, the designer can move an avatar around the plane and collaborators, and get an avatar-centred view on the 3-D environment (Figure 7, left). Cones are used to simplistically represent the cameras (red cones) and the avatar (pink cone).

Finally, according to the avatar's position in the environment, a PDA emulation displays the avatar's coordinates, the available collaborators' names, even those who are not in eye contact, the activated services (WiFi, camera detection), and the camera's video feedback that used to indicate that the avatar is visible (pink cone visible) or where an area of localisation exists. This simulation provides the designer with a mean to evaluate the adequacy of:

- The content of the information provided to a future user with the task considered
- The data representation with the future user's expertise: label and/or video-feedback can be used to represent the availability of the video-based localisation service
- The position of the camera required for any localisation-based service: if the designer notices that the avatar is always in places where services are unavailable, the designer will be encouraged to modify the initial design of the considered mobile mixed system. For example, solutions may be the definition of a

Figure 7. SIMBA windows when simulating the mobile mixed aeronautical maintenance situation



different disposition of the physical space or, the use of different or additional resources (e.g., adding more camera to support the video-based localisation or replacing it with an RF-ID based localisation).

Modifying elements of the ASUR model of the mobile mixed systems allows the testing of different localisation techniques, content, and representations of the dialog. Following such modification, SIMBA elements clearly associated with the modified ASUR elements must then be adapted while the other SIMBA elements remain unchanged. Based on this simulation environment, advantages and limitations of video-based localisation techniques vs. an RF-ID based system for example can be assessed in flexible manner.

Through this example, we have illustrated that our three requirements are fulfilled:

- The simulation is relying on an ASUR design model of the system
- SIMBA elements of the simulation have been directly reused from previous use of the platform: wall, humans, plane, camera-based localisation
- SIMBA elements have been specifically developed for entities used in this interaction situation: PDA emulation, application specific functions

Practice and Experience Feedback of the Use of SIMBA

Five different persons have been using SIMBA to build an ASUR-based simulation. Informal interviews were conducted to retrieve their comments.

SIMBA 3-D scene management and easy insertion of elements in the digital environment through the 2-D map are the two major advantages of SIMBA. The mechanism in charge of the automatic event forwarding to relevant entities, according to the ASUR model, is also really helpful: ASUR adds an abstract level for reasoning and identifying data exchanges to the simulation environment. Moreover, SIMBA elements model enforces to decouple presentation and functional aspects of the application. Finally, the predefined methods of the functionality component are well suited to the needs of the simulation.

However, when starting the simulation, the creation and loading order of the different SIMBA elements has a very big impact on the simulated application. This clearly appears to be restrictive in some cases. For example, it is not straightforward to create a generic PDA because its emulation depends on the data that will be displayed. Other identified drawbacks are the heavy link with Java3D and the lack of a clear documentation.

Regarding the outcomes in terms of evaluation of the simulated systems, moving the avatar and understanding the relationship between output devices in the 3-D scene and their emulation displaying the information appears to be less disturbing than we thought: the emulation window is considered as a tool-palette, which is a usual way for presenting options in interactive systems. Users of the simulation were thus not disturbed by this aspect. In addition, playing with the avatar gives a better representation of the scenario that has to be performed. Alternatives would have required paper prototypes or 2-D interfaces, which would have cut them from the physical environment.

The main drawback reported is that the user does not feel completely immersed in the simulated mixed interaction space. In fact, we do not consider this as a negative aspect on our approach. We are not chasing a completely realistic environment for the simulated system, but rather an environment that supports the analysis of aspects specific to mobile systems: localisation techniques, presentation mode, impact on the context or other persons, and so forth. Finally, interacting through the keyboard is tiresome: a wider range of input devices should be provided according to the 3-D situations and the skills of the users.

As a result, SIMBA is based on outcomes of early design phases, and constitutes an important first step towards the simulation of mobile mixed systems in order to prototype and evaluate such systems.

FUTURE TRENDS

To build upon this coupling of design results and implementation process, improvements to the simulating environment and the model on which it is based are already envisioned.

In Terms of the ASUR Model

We have illustrated that ASUR now covers the design of every kind of mixed systems, from augmented reality to augmented virtuality. But ASUR only provides a part of the description of the user's interaction with a mixed system: it highlights domain concepts, partially represents presentation and dialog aspects, but does not really cover task sequences, system and platform models, device models, and so forth. Linking this model with complementary tools or notations would lead to a more complete modelling of the user's interaction with a mobile mixed system. These additional aspects could then be taken into account into the simulation step, thus enhancing the insight of the simulation. For example, coupling ASUR with a task-based approach is already ongoing, and would help a designer in analysing the impact on the user of changing from one mixed interaction technique to another one.

To support the articulation of these different models, we believe that adopting a model driven engineering (MDE) approach will greatly enhance and support a more systematic design method. MDE approach is also useful to transform an abstract model into a more concrete model such as software component description, for example (Dubois, Gauffre, Bach, Salembier, 2006).

In Terms of SIMBA

We ran different simulations of our application scenario with SIMBA, using various sensing technologies. The required SIMBA elements have thus been added to the SIMBA library. However, more elements must be developed to a sufficient basis for the rapid simulation of mobile mixed systems.

Technically, some improvements are required to better address the creation process of the different elements when starting a simulation, and to enable the use and development of more generic elements. To do so, technologies such as runtime loading of classes and CorbaCCM are envisioned.

We also reported, in this chapter, results of a limited evaluation. A more detailed experiment, based on the same scenario, is planned for this year. Real operators and engineers will be involved so we can learn, from end-users, the benefit of this simulation.

In terms of use of the platform to prepare the simulation, some effort is required to support, in a more usable way, the definition of characteristics not listed in the ASUR models, properties of the 3-D environment, and so forth.

Finally, our ultimate goal is to be able to substitute part of the simulated interaction by their tangible counterparts: the simulation of the mixed system would then gradually change into a mixed system instead of a virtual application. Indeed, one can imagine that the PDA does not have to be simulated and can be directly handled by the user of the SIMBA simulation; manipulating a physical wand to point at a given place on the screen can also be extracted from the simulation, and so forth. This would support an incremental prototyping approach of mobile mixed systems.

CONCLUSION

This chapter introduced an extension of ASUR, a mixed system model, and a platform to simulate mobile mixed systems previously modelled with ASUR. The extended version of ASUR provides designers with a tool to reason about entities, different interaction facets, and data flows involved in mixed systems. The extension is based on established results in ergonomics: predictive analysis of ergonomic properties based on the model is thus supported.

Based on this modelling approach, SIMBA constitutes a support for running a simulation of an ASUR-modelled mobile mixed system:

preexisting SIMBA elements ready to use in a simulation can be very simply associated to elements of the ASUR model, so that the simulation of the mobile mixed system becomes independent of the programmer's skills to manage the available virtual reality platforms, 3-D toolkits, and authoring systems. For more experienced users, SIMBA also offers a predefined format for creating additional SIMBA elements.

This form of prototyping clearly complements the predictive analysis performed on the model by tackling technological, topographical, and dynamic aspects of the system that are not easily visible on a modelling approach or traditional low-fidelity prototyping approaches.

Combining these two approaches thus constitutes a first step toward the tight coupling of designing and implementing steps of a mobile mixed system development. As highlighted in this chapter, interlacing a model-based approach and the prototyping approach only covers two steps of the traditional participatory design approach. Further work will lead to articulate tools and methods in order to equip the whole iterative cycle: improving the SIMBA architecture model will allow its connection to other component platforms; usability property expression in terms of ASUR and SIMBA will support the evaluation of mobile mixed systems specific aspects. This work is thus a first step towards a well-founded set of models, well integrated with one another, and that fit into a development process of mobile mixed systems.

ACKNOWLEDGMENT

Part of this work has been financially supported by the French National Research Center (Grant CNRS–GDR-I3–AS153) and the Région Midi-Pyrénées (Grant for the project Protopraxis). Authors would like to thank Isabelle Fonquernie for help improving SIMBA and Amélie Rault-Azeem for her thorough proofreading.

REFERENCES

Azuma, R. T. (1997). A survey of augmented reality. *Presence: Teleoperators and Virtual Environments, 6*(4), 355-385.

Bach, C., & Scapin, D. L. (2003). Adaptation of ergonomic criteria to human-virtual environments interactions. In *Proceedings of Interact'03* (pp.880-883). Amsterdam: IOS Press.

Balcisoy, S., Kallmann, M., Fua, P., & Thalmann, D. (2000). A framework for rapid evaluation of prototypes with augmented reality. In *Proceedings of VRST'00* (pp. 61-66). NY: ACM Press.

Bastien, J. M. C., Scapin, D. L. (1993). *Ergonomic criteria for the evaluation of human-computer interfaces* (Tech. Rep. No. 156). Rocquencourt, France: INRIA.

Bauer, M., Bruegge, B., Klinker, G., MacWilliams, A., Reicher, T., Riß, S., Sandor, C., & Wagner, M. (2001). Design of a component-based augmented reality framework. In IEEE (Ed.), *Proceedings of ISAR01* (pp.45-54).

Benard, R., Chevalier, P., & Baudoin C. (2004). Simulation participative et immersive.Retrieved January 12, 2006, from http://www.cerv.fr/fr/activites/SPI.php

Caubet, R., & Torguet, P. (1995). Viper: A virtual reality applications design platform. In *Proceedings of Eurographics Workshop on Virtual Environments.* Monaco.

Coutrix, C., & Nigay, L. (2006). Mixed reality: A model of mixed interaction. In *Proceedings of AVI06* (pp.45-53). NY: ACM Press.

Cruz-Neira, C, Bierbaum, A., Hartling, P., Just, C., & Meinert, K. (2002). *VR Juggler - An open source platform for virtual reality applications.* Paper presented at the AIAA Aerospace Sciences Meeting and Exhibit, Reno, NV.

Delotte, O., David, B., & Chalon, R. (2004). Task modelling for capillary collaborative systems based on scenarios. In *Proceedings of TAMODIA'04* (pp.25-31). NY: ACM Press.

Dubois, E., Gauffre, G., Bach, C., & Salembier, P. (2006). Participatory design meets mixed reality design models - Implementation based on a formal instrumentation of an informal design approach. In *Proceedings of CADUI'06* (pp.71-84). Berlin: Springer-Verlag.

Dubois, E., Gray, P. D., & Nigay, L. (2003). ASUR++: A design notation for mobile mixed systems. *Interacting with Computers, 15*(3), 497-520.

Dubois, E., Nigay, L., Troccaz, J., Chavanon, O., & Carrat, L. (1999). Classification space for augmented surgery, an augmented reality case study. In *Proceedings of Interact'99* (pp. 353--359). Amsterdam: IOS Press.

Dupuy-Chessa, S., & Dubois, E. (2005). Requirements and impacts of model driven engineering on mixed systems design. In *Proceedings of IDM'05* (pp.43-65), Paris, France.

Feiner, S., MacIntyre, B., & Seligmann, D. (1993). Knowledge-based augmented reality. *Communication of the ACM*, *36*(7), 53-61.

GUIDE-Me. Retrieved from http://liihs.irit.fr/guideme

Haller, M., Zauner, J., Hartmann, W., & Luckeneder, T. (2003). *A generic framework for a training application based on Mixed Reality* (Tech. Rep). Upper Austria University of Applied Sciences, Vienna.

Hilliges, O., Sandor, C., & Klinker, G. (2005). *Interaction management for ubiquitous augmented reality user interfaces.* Diploma Thesis, Technische Universität München.

Hong, I. J., & Landay, J. A. (2000). SATIN: A toolkit for informal ink-based applications. In the *Proceedings of UIST'00* (pp.63-72). NY: ACM Press.

Ishii, H., & Ullmer, B. (2000). Emerging frameworks for tangible user interfaces. *IBM Systems Journal, 39*(3/4), 915-931.

Kähäri, M., & Murphy, D. J. (2006). MARA – Sensor based augmented reality system for mobile imaging. In *Proceedings of ISMAR'06*. NY: ACM Press.

Kato, H., & Billinghurst, M. (1999). *Marker tracking and HMD calibration for a video-based augmented reality conferencing system*. Paper presented at IWAR'99, San Francisco, CA.

Leulier, C., Bastien, J. M. C., & Scapin, D. L. (1998). *Compilation of ergonomic guidelines for the design and evaluation of Web sites* (Commerce & Interaction Report). Rocquencourt, France: INRIA.

Mackay, W. E. (1996). Réalité Augmentée : le Meilleur des Deux Mondes. *La Recherche (285)*, 80-84.

Mackay, W. E., Ratzer, A., & Janecek, P. (2000). Video artifacts for design: Bridging the gap between abstraction and detail. In *Proceedings of DIS2000*. NY: ACM Press.

Milgram, P., & Kishino, F. (1994). A taxonomy of mixed reality visual displays. *IEICE Transactions on Information Systems, E77-D* (12), 1321-1329.

Molina Masso, J. P., Vanderdonckt, J., Gonzalez Lopez, P., Caballero, A. F., & Lozano Perez, M. D. (2006). Rapid prototyping of distributed user interfaces. In *Proceedings of CADUI'06* (pp.153-168). Berlin: Springer-Verlag.

Noma, H., Miyasato, T., & Kishino, F. (1996). A palmtop display for dextrous manipulation with haptic sensation. In *Proceedings of Conference on Human Factors in computing systems* (pp.126-133). NY: ACM Press.

Norman, D. (1986). *Cognitive engineering in user centered system design: New perspectives on human-computer interaction*. Lawrence Erlbaum Associates.

Renevier, P. (2004). *Systèmes Mixtes Collaboratifs sur Supports Mobiles: Conception et Réalisation*. Unpublished doctoral dissertation, University Joseph Fourier, France.

Renevier, P., Nigay, L., Bouchet, J., & Pasqualetti, L. (2004). Generic interaction techniques for mobile collaborative mixed systems. In *Proceedings of CADUI'04* (pp.307-320). NY: ACM.

Rodden, T., Cheverst, K., Davies, N., & Dix, A. (1998). Exploiting context in HCI design for mobile systems. In C. Johnson (Ed.) *Proceedings of the First Workshop on Human Computer Interaction with Mobile Devices.* University of Glasgow.

Rodriguez, N., Jessel, J. P., & Torguet, P. (2001). Asset: A testbed for teleoperation systems. In *Proceedings of SIM01.* WSES-IEEE.

Sanchez-Segura, M. (2005). *Developing future interactive systems.* Hershey, PA: Idea Group Publishing.

Shaer, O., Leland, N., Calvillo-Gamez, E. H., & Jacob R. J. K. (2004). The TAC paradigm: Specifying tangible user interfaces. *Personal and Ubiquitous Computing, 8*(5), 359-369.

Tramberend, H. (1999). Avocado: A distributed virtual reality framework. In *Proceedings of VR'99* (pp.75-81). Washington DC: IEEE Computer Society.

Trevisan, D. G., Vanderdonckt, J., & Macq B. (2005). Conceptualising mixed spaces of interaction for designing continuous interaction. *Virtual Reality, 8*(2), 83-95.

Ullmer, B., & Ishii, H. (1999). MediaBlocks: Tangible interfaces for online media. In *Proceedings of CHI'99,* vol. 2 (pp.31-32). NY: ACM Press.

KEY TERMS

Aeronautical Maintenance: Activity that consists in preserving planes from failure, decline, or accident. Due to the amount of documentation, space to observe, and operator's constraints, mobile mixed systems constitute promising solutions to support the aeronautical maintenance.

ASUR Model: Design model describing the entities involved in the user's interaction with a mixed system.

Component Model: Structure adopted to develop any software element of a given application or library.

Mixed Systems: Interactive systems involving physical and digital entities. It covers interactive systems in which digital objects enrich the user's interaction with physical objects and interactive systems in which physical objects are manipulated to support the user's interaction with a digital application. Other terms are also used to cover part of mixed systems: augmented reality, tangible user interfaces, and mixed reality systems.

Mobile Mixed Systems: Interactive systems supporting the user's mobility and making extensive use of the physical nature of the user's environment.

Model-Based Design: Early step of a development process that consists in describing the object of the design (software, interaction techniques, etc.) with a formalised notation. The notation can be textual, graphical, or a combination of both. A notation is formalised if it conforms to a metamodel.

Model-Driven - Simulation-Based Prototyping (MD-SBP): Simulation-based prototyping of an interactive application tightly coupled with the description of the application using a given model.

SIMBA: Platform for model-driven – simulation-based prototyping of mobile mixed systems. This platform, based on the ASUR model, provides extension facilities to add new simulated elements.

Simulation-Based Prototyping (SBP): Design step that aims at producing a functional form of an interactive application that can be manipulated through a 3-D virtual environment. Cheaper and easier to produce than a high-fidelity prototype, a simulation-based prototype includes the dynamic aspects of the interactive application and the physical aspects and constraints of the interactive situation.

Chapter XXIII
Engineering Emergent Ecologies of Interacting Artefacts

Ioannis D. Zaharakis
Computer Technology Institute, Greece

Achilles D. Kameas
Computer Technology Institute, Greece
Hellenic Open University, Greece

ABSTRACT

Nowadays, our living environments already provide ubiquitous network connectivity and are populated by an increasing number of artefacts (objects enhanced with sensing, computation, and networking abilities). In addition, people are increasingly using mobile devices as intermediaries between themselves and the artefacts. In order to create, manage, communicate with, and reason about ubiquitous computing environments that involve hundreds of interacting artefacts and cooperating mobile devices, we propose to embed, in these entities, social memory, enhanced context memory, and shared experiences. In this context, we describe an engineering approach and a framework to deal with emergent ecologies of locally interacting artefacts that provide services not existing initially in the individuals, and exhibiting them in a consistent and fault-tolerant way. Because they are emergent, their structure or availability are not predefined or known before hand; we draw from swarm intelligence methods to describe such ecologies.

INTRODUCTION

Already, an increasing number of sensors are becoming embedded in the everyday objects or in the environment at a low cost. As a result of this continuing trend, an elementary ambient intelligence (AmI) infrastructure has become installed (though still fragmented), information appliances are commercially available, and ubiquitous computing (UbiComp) applications (currently in the form of games and informative services) are being deployed. As Norman (1998) anticipated, with the proliferation of networks, information appliances, and artefacts, large amounts of data start being diffused in our living environment, and knowledge about patterns and context of human activities are generated. In addition, new generations of mobile devices (such as mobile phones, tablet PCs, PDAs) are being developed having increased capabilities and resources. These devices can now be considered as powerful information processing, storage, and access tools that can be used as facilitators between people and a smart environment, as they can be aware of the artefacts in their vicinity (Lopez de Ipina, Vaszquez, Garcia, Fernandez, & Garcia, 2005).

These developments have the potential to greatly enhance human activities (i.e., by automating dull or ordinary tasks; by speeding up time to exchange data, records, and files; by providing ubiquitous access to services and infrastructures, etc.). We need, however, to overcome the current limitations of distributed and context sensitive computing basically due to the classical client-server approach that is embodied in most, if not all, object-oriented environments and move towards reusable components and peer-to-peer (P2P) networks. By encouraging the adoption of agents and services as a building block for future advanced applications, full delegation of tasks in societies of interconnected computational and human resources can be achieved.

As a consequence of the availability of new technologies, the nature of the human activities eventually assisted by artefacts is rapidly changing. People have to (consent to) build new task models or adapt the ones they have already been using, a task not trivial at all. Execution of new tasks may become difficult due to the inherent systemic complexity of UbiComp applications that, among others, result from device incompatibility, and a huge number of interactions among visible and nonvisible actors. As ambient intelligence becomes widespread, people with low levels of IT literacy will be increasingly asked to interact with smart objects. Humans, with their "analogue" way of thinking and acting, have difficulties in using digital systems, because the latter demand precision, cannot tolerate misuse, and are unable to adapt to changes in operating environments (Norman, 1998). In addition, people will have to adjust to task execution involving high degrees of interruption and task switching. This situation might lead to the social exclusion of those not able to cope with this complexity, and to possible failure of realizing the AmI vision (ISTAG, 2001). Mobile devices are expected to play an important role in the adoption of ambient intelligence because we have already become familiar with using them, albeit in a simpler context.

This work builds upon the envisaged structure of the AmI environment as one populated by thousands of communicating tangible objects and virtual entities (Kameas, Bellis, Mavrommati, Delaney, Colley, & Pounds-Cornish, 2003). Following an agent-oriented approach and adopting principles of Nouvelle AI (an alternative to the symbolic representation of internal models of the world, promoting that intelligence, as expressed by complex behaviour, "emerges" from the interaction of a few simple behaviours), the next sections describe a conceptual framework, which has been inspired by biological structures and is capable of dealing with phenomena emerging in such an environment. Finally, an engineering approach related to new research issues and requirements is introduced.

On the road to realising UbiComp applications and AmI spaces, several technical issues need to be resolved in order to make these systems adoptable and usable. Some of the major requirements a UbiComp system has to confront are: mask the *heterogeneity* of networks, hardware, operating systems, and so forth; tackle *mobility* and *unavailability* of nodes; support component *composition* into applications; *context awareness*; preserve object *autonomy* even for *resource constraint* devices; be *robust*, *fault tolerant*, and *scalable; adapt* to environmental changes; and *be usable by novice users* via understandable designed models (Drossos, Mavrommati, & Kameas, 2007). The approach followed by current efforts assumes a network infrastructure that allows direct communication of application components, that is, UbiComp systems are largely treated as (a) distributed systems with resource constrained nodes,

(b) software components that interact using message exchange or parameter passing, or (c) mobile ad-hoc networks whereby some nodes may fail or be unavailable. Within such systems, objects have computerized interfaces (i.e., containing screens, keyboards, command buttons etc.), which make them much different from the objects that populate our everyday spaces.

For example, in Humble *et al.* (Humble, Crabtree, Hemmings, Akesson, Koleva, Rodden, & Hansson, 2003), the "jigsaw puzzle" metaphor is adopted in the interface. Objects are represented as puzzle-piece-like icons that the user "snaps" together to build an application. While this metaphor is comprehensible, the interactions are simplified to sequential execution of actions and reactions depending on local properties (e.g., sensor events), which limits the potential to express many of the user's ideas. In Truong et al. (Truong, Huang, & Abowd, 2004), a pseudonatural language interface based on the fridge magnet metaphor is proposed, while in the browser approach of Speakeasy (Edwards, Newman, Sedivy, Smith, & Izadi, 2002), components are connected using a visual editor based on file-system browsers. ZUMA (Baker, Markovsky, van Greunen, Rabaey, Wawrzynek, & Wolisz, 2006) is a platform, based on a set of clean abstractions for users, content, and devices, that supports the configuration and organization of content and networked heterogeneous devices in a smart-home environment. ZUMA employs the notion of multiuser optimal experience and attempts to achieve optimization by migrating applications to different environments. However, the system operates using fixed rules, without taking into consideration user's feedback.

In this chapter, we intend to present a different conceptualization that is based on the way natural, living systems use local interactions to self-organize and behave coherently. In our framework, the environment is used as the communication medium; objects need not be aware of each other, and self-organization results from a multitude of local interactions. People are part of this ecology, using objects to perform tasks, only being aware of object affordances. Thus, there need not exist a computer-like screen-based interface for the ecology; instead, the ecology exchanges information with people using the most suitable object (suitability depends on context) in their environment.

BACKGROUND

An AmI space is a hybrid (i.e., physical and digital) environment populated by a large number of communicating tangible objects and virtual entities. An AmI space provides infrastructure to support the composition, deployment, and usage of distributed applications and services. As computational power is diffused in our living/working environment, and the number of everyday devices that are capable of sensing, processing, and communicating continuously grows rapidly, new requirements are posed by the heterogeneity of the involved devices, the complexity of interactions, and the need for simple usage models.

Although the constituent components may have restricted resources, the huge number of interactions gives rise to emergent phenomena. Thus, new research issues arise concerning i) the system complexity that results from the thousands of local interactions and their effect on system stability; ii) the need for flexible and dynamic system architecture capable to evolve and adapt to new situations and configurations; iii) the context dependence of system operation; and iv) the human involvement, which calls for new, more natural, human-machine interaction schemes.

The study of emergence as an inherent property of AmI spaces, as the complexity of AmI systems increases, is of paramount importance if one wants to be able to regulate their behaviour. The proliferation of embedded computers (i.e., in mobile phones, CCTVs, PDAs, smart appliances, etc.) and the huge increase in the available network bandwidth will lead to the development of large computation platforms (i.e., "global computers") capable of supporting complex applications. Artificial intelligence (AI) may prove the decisive success factor by contributing techniques to help us deal with research issues at all levels, including bio-inspired computation models, services aimed at end users, as well as machine to machine services, dynamic composition and adaptability, context awareness, autonomy, and semantic interoperability.

Bio-Inspired Approaches, Complexity, and Emergent Behaviours

When studying natural systems consisting of many living organisms, properties such as stability, coherence, flexibility, and adaptability can be observed. These natural organisms exhibit these characteristics because they are integrated and optimized with respect to their computation and control strategies, morphology, materials, and their environment (Knoll & de Kamps, 2004). Moreover, the capabilities of the ecology are distributed over the whole system, and the physical phenomena are created by the interaction of the participants with their environment. Examples of such natural systems include insect colonies, flocks of birds, schools of fishes, herds of mammals, and so forth. Interestingly, these systems demonstrate coherent collective behaviours, although there is no evidence of some kind of centralised control or leadership.

The ICT systems that exhibit global complex behaviour emergent by the local interactions of their simple components are referred to as complex systems (Bullock & Cliff, 2004). Depending on the analysis level, these systems can be described either as simple components that interact with each other in relatively simple ways, influencing only their neighbours, or as a system exhibiting a complex overall behaviour. Thus, when engineering an artificial complex system to operate "intelligently," the increased task complexity can experience significant reductions if each component in the control loop solves a simplified problem while relying on the other components to create the conditions that make these simplifying assumptions valid (Brooks, 2002).

In current AI research, there is an increasing interest in how to engineer autonomous entities with limited capabilities in both peripheral (sensors, actuators) and computational resources (processors, memory, communication, etc.), yet simultaneously, exhibiting robustness and behavioural agility. The swarm intelligence field contributes to this effort by focusing on the emergent collective intelligence of (unsophisticated) agents that interact locally with their environment (Bonabeau, Dorigo, & Theraulaz, 1999; Kennedy & Eberhart, 2001). These unsophisticated agents are referred to, in the literature, as simple reflex agents (Russell & Norvig, 2003) or purely reactive agents (Wooldridge, 1999), thought the latter are more hardware oriented.

Dealing with Symbiotic AmI Spaces

Based on the above-mentioned analysis and focusing on the creation, management, communication with, and reasoning about UbiComp environments that involve hundreds of interacting and cooperating devices, we propose a bio-inspired engineering approach, and a framework to deal with emergent ecologies of locally interacting artefacts with computing and effecting capabilities that provide services not existing initially in the individuals, and exhibiting them in a consistent and fault-tolerant way. Furthermore, in order to reduce the difficulty of carrying out everyday activities in an AmI environment, we consider the delegation of certain tasks to a digital alter-ego (roughly related to ISTAG, 2001) that continuously observes user's interactions with digital objects in different contexts, can learn the user's interests and habits, and can evolve to take initiative in well-known and harmless activities. This system may have a wide family of forms and means to serve the user; it should pervade the living space, and manifest itself unobtrusively by using the essential available artefacts. As delegation becomes the key issue, it is evident that an agent-oriented or service-oriented view, as recently described, among others, by Foster *et al.* (Foster, Jennings, & Kesselman, 2004), are the most suitable.

Interaction

Within such an AmI space, people still have to realize their tasks, ranging from mundane everyday tasks (i.e., studying, cooking, etc.) to leisure or work-related tasks, or even tasks that relate to emergency situations (i.e., home care, accident, unexpected guests, etc.). To do so, they have at their disposal the objects that surround them. These, in fact, are new or improved versions of existing objects that, by using information and communication technology (ICT) components (i.e., sensors, actuators, processor, memory, wireless communication modules), can receive, store, process, and transmit information, thus allowing people to carry out new tasks or old tasks in new and better ways. In the following, we shall use the term "artefacts" for this type of augmented objects. Turning an object into an artefact is a process that aims at enhancing its characteristics, properties, and abilities so that the new affordances will emerge. In practical terms, it is about embedding in the object the necessary hardware and software modules.

From the interaction point of view, we are mostly concerned with the interface of the artefacts and the collective interface of UbiComp applications. The former shall directly affect or depend upon the physical form and shape of the artefacts. The latter can exist in the "digital space" of a computer, that is, a PDA that runs specific software representations of the artefact services.

Thus, people interact with an AmI environment in order to (Kameas, Mavrommati, & Markopoulos, 2004):

- **Engineer a UbiComp application within the environment**, as a composition of artefacts that collectively serve a specific purpose or satisfy a declared set of needs.
- **Use an application to satisfy their needs:** Such an application may be composed by people themselves, or could be bought and installed.

Interaction takes place in two levels:

- **Artefact-to-artefact:** The objects themselves may form an "underlying" layer of interactions, mainly in order to exchange data and to serve their purpose better. Such interactions may use wired channels or any of the available wireless protocols (in a peer-to-peer or broadcast manner), or even the Internet.
- **User-to-application:** The user interacts a) with any single artefact b) with a collection of cooperating devices. Moreover, one has to consider the case where many users interact with the same application.

The degree of visibility and control that people may have on these interactions may vary depending on people's ability to perceive the system state: any of these two types of interaction may happen either explicitly or implicitly.

An explicit interaction that happens under the control of people always provides feedback about its state to them. Although this may seem desirable, it may also become very annoying if one takes into account that there will be hundreds of artefacts in our environment. People interact explicitly with objects or services (or collections of these). In the case of individual objects (or, preferably, services), interaction can be supported by the object affordances. When people interact with a UbiComp application composed of a collection of objects, its "affordances" have to emerge and be made explicit.

Implicit interactions are usually under the control of actors other than people; these could be processes, artefacts, intelligent agent mechanisms, or even artefact owners. Implicit interactions can only be acceptable if they can be trusted, and do not violate privacy or ethics. People need not be directly aware of the communication among objects. Moreover, even certain interactions with UbiComp applications should happen unobtrusively, that is, people should be made aware of the state changes of the application components without being disrupted from their current tasks.

Finally, one should also consider the context of interaction, which ranges from public to private (with respect to disclosure), from individual to shared (with respect to stakeholders) and from closed to open (with respect to space).

A symbiotic AmI space, as described in the next section, should enable people to simply act upon its objects to use their services; no special collective interface should be required.

CONCEPTUAL FRAMEWORK

When considering AmI spaces, such as those described (e.g., an everyday living or working space), populated by many actors with digital selves (artefacts and human), one has to deal with undoubtedly complex emergent phenomena (Lindwer, Marculescu, Basten, Zimmermann, Marculescu, Jung, & Cantatore, 2003). By viewing the user as part of the system, one has to ensure an unobtrusive symbiosis between human and artificial entities, and to establish the user acceptance and confidence towards the UbiComp applications. Our approach is based on a human-centric autonomic system (Mitchell Waldrop, 2003) constituting of self-managing ecologies that are diffused in the everyday living space. In such ecologies, artificial entities coexist unobtrusively with humans, and perform collaborative tasks through a continuous evolvable process concerning both their physical and social cognitive growth (we call them "*ambient spheres*").

It is apparent that the social dimension is a significant factor that characterises such systems. Technically, the reproduction of social behaviours and the handling of complex tasks with an equal agility, as the one exhibited by natural intelligent systems, can be achieved by i) considering that all the necessary information lies out in the environment and surrounds the participants (according to Brooks, 1991, 2002), and ii) using bio-inspired approaches in designing intelligent systems, in which autonomy, emergence, and distributed functioning are promoted (Bonabeau *et al.*, 1999; Kennedy & Eberhart, 2001). By distributing the individual physical/computational/cognitive capabilities over the entire ecology, and by immersing the ecology into a UbiComp environment, we can generate theory and technology for the understanding of the own self and its relation with the surrounding world. In order to deal with the collective behaviour of large societies in situated domains, the system has to perform analysis and synthesis of primitive behaviours that result from individual interactions. This is originated by the engineering methodology proposed by Brooks (1991) (and followed by Mataric, 1995 in the multiagent domain), who decomposes the system into parts, builds the parts, and then interfaces them into a complete system. The decomposition is by activity, and the advantage of this approach is that it gives an incremental path from very simple systems to complex autonomous intelligent systems. At each step of the way, it is only necessary to build one small piece and interface it to an existing, working, complete intelligence. We extend the above consideration by additionally i) attributing AmI objects with physical expression (dimensions, shape, texture, colour, plugs, sockets, connectors, etc.), and ii) dealing with the provided services as basic behavioural building blocks of the overall system behaviour.

According to our approach, a living/working AmI space is populated with many heterogeneous objects with different capabilities and provided services. All these objects and services are regarded as basic building blocks, having an internal part that encapsulates the internal structure and functionality, and an external part that manifests the capabilities and influences the surroundings. Additionally, every basic building block has several predefined functions (we call them *basic behaviours*). Some of the basic behaviours are just reactions to external events, and some are continuously pursued to be fulfilled. The former type of basic behaviours imitates the reflex actions of the living organisms, while the latter the preservation instincts. The interrelationships between the basic building blocks and the associated environment form an *ecology*. Definitely, until now, our approach does not differ a lot from a natural ecosystem, or from (purely reactive) collective robotics, except that we consider services as part of the swarm, too. However, we use another ingredient, called *ambient system* (AmS), with the following characteristics:

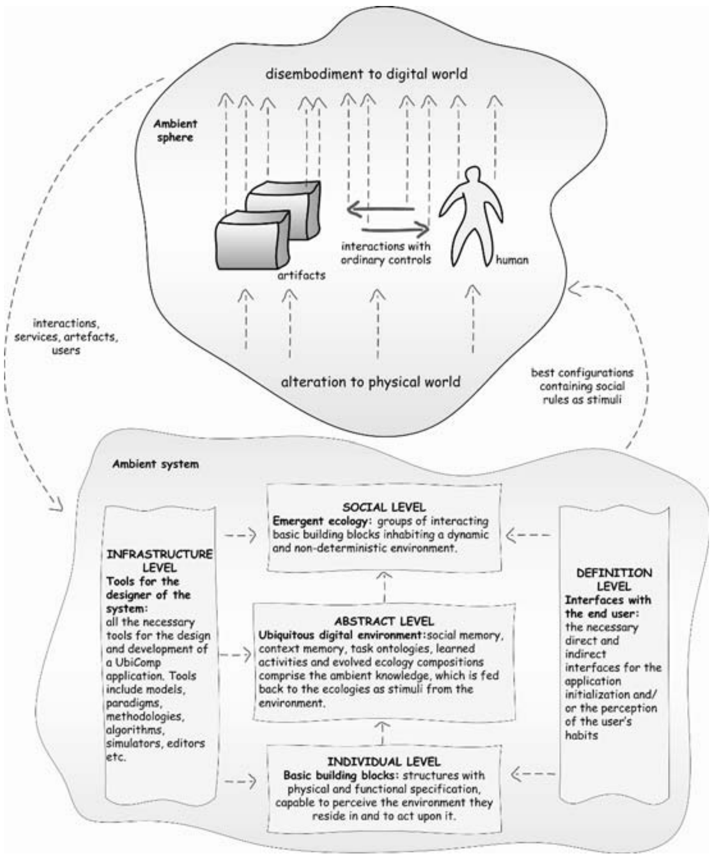
- It acts as the "glue" between the tangible and nontangible basic building blocks of the ecology by providing an interface definition language (IDL) and thus, integrating the basic building blocks into a common interoperability framework. A similar approach is adopted by the common component architecture (Larson, Norris, Ong, Bernholdt, Drake, El Wasif, *et al*., 2004) (CCA) that uses the Babel (Dahlgren, Epperly, & Kumfert, 2003) language to allow components written in various languages to interoperate. One step to this end is the design and implementation of a hierarchy of multidimensional ontologies that will include both nonfunctional descriptions, and rules and constraints of application, as well as aspects of dynamic behaviour and interactions. A core ontology will be open and universally available and accessible. During the ecology lifetime, the core ontology will be supplemented with higher-level goal, application, and context specific ontologies. These ontologies, describing specific application domains, can be proprietary. Emerging behaviour, in this context, will be considered as a result of interactions among heterogeneous, seemingly incompatible or non-predefined entities. Moreover, all higher-level constructs will be inherently able to use all the knowledge they will be able to access.

- It is responsible for the observation and collection of the interactions between AmI objects, provided services, and users. The collected data are used to create best-practice ecology configurations that will help the gradual accumulation of social memory. Aspects of social intelligence are embodied with the ecology configurations as basic social behaviours aiming to regulate the group interactions. These social behaviours are provided as ontological constructions that are also subject to evolution.
- It uses the collected information as input to appropriate reinforcement learning algorithms (e.g., Q learning algorithm, (Watkins & Dayan, 1992)) in order to learn the suitable configurations in association to the task to be accomplished. On the other hand, it utilises a genetic algorithm responsible for the ecology evolution in terms of resource availability. To this end, the modelling of "user perceived quality" (or user comfort (Mozer, 2005)) in the AmI space is necessary, and this could be a parameter of the fitness (evaluation) functions that could be used in the evolution and learning processes.
- It provides feedback to the members of the ecology (social memory in association to "user perceived quality") favouring the best configurations and implicitly assigning cognition to the whole ecology. It is mentioned that the feedback information is provided to the ecology as another stimulus or stimuli and thus, does not require extra (complicated) sensors mounted to the artefacts. This would require a different modelling and engineering approach comparatively to the one described so far. Instead, and along the lines of swarm intelligence, where the environment is a stimulus for the swarm, we treat the AmS as another (special) basic building block; the environment building block.

The AmS is realised as a distributed platform that supports the instantiation of ambient spheres, each of which are formed to support human activity (Figure 1). Examples of candidate AmS technologies are the currently available distributed component frameworks and service-oriented architectures. An ambient sphere is an integrated autonomous system realised as a set of configurations between the AmI objects and the provided services into the AmI environment. In our model of discourse, the end-users are placed inside the ambient sphere, as this allows us to model them as another basic building block that generates events and changes the environment. Then the sphere can evolve on its own (in a sense, it develops a "self," as it becomes aware of its capabilities and the context of operation, while it maintains a set of goals to achieve), by (self-) configuring and (self-) adapting to better satisfy certain user needs.

Summarising, the ambient system, which includes networking, middleware, learning, and evolution algorithms and mechanisms, ontologies describing structural and (both individual and

Figure 1. Individual and Social levels correspond to the basic building blocks and ecologies, respectively. Abstract level encloses the social memory of the ecologies; such knowledge must be transferred to the ecologies implicitly, for example, as stimuli of the environment, since individuals and, consequently, the emergent ecologies do not contain any knowledge representation scheme nor reasoning mechanism. Infrastructure level provides system designers with the appropriate tools to develop a system. Definition level is the user interface with the final user.



social) behavioural aspect, and so forth, provides an integration framework to collaborating objects and services in order to support user activities by forming ambient spheres. The driving force behind the sphere formation is the selfishness of objects and services; they try to "survive" by operating within a given set of resources. The antagonism of the members is reflected on the spheres, which are formed in an ad-hoc (emergent) way, and they persistently try to meet the "success criterion": user satisfaction.

There are many potential benefits of such an approach including greater flexibility and adaptability of the system to the environment, robustness to failures, and so forth. The swarm comprises different typologies of ecologies; thus it is heterogeneous, also, from the point of view of the provided services. Such differences contribute to the overall capabilities of the system. As a general principle, the services are as simple as possible. The swarm system aggregates the capabilities of extremely simple members by increasing the number of agents, supporting the sharing of resources and maximizing the effectiveness of communication. Ideally, the composition emerges based on previous interactions, and on the context (time and place) they took place. The proposed framework adopts an innovative approach in that:

- It applies to software and services domain an approach previously applied only to robotic applications (which perform local decision making, are inherently distributed, and consist of interacting parts)
- It constitutes an attempt to build integrated UbiComp applications from the services of an AmI environment (without modifying the services, but by using the environment as the integration medium)
- It deals with different and heterogeneous actors (objects, services, social rules, spheres, and people) by applying a hierarchy of semantically-rich representations

Together with the advantages and the benefits of the proposed framework goes a series of technological issues that entail engineering and evaluation. The most important of them are considered in the following section.

An Example: Virtual Residence

As we move from physical to digitized spaces, some of the existing real-world components, concepts, or metaphors will have to be adapted. One important component of human life is residence. It does not only represent the space where one lives, but also a sphere that encompasses one's activities, a place where one can seek refuge, a repository of one's objects and experiences, a private space where more intimate interactions can take place, and many more. It is for these reasons that people feel uncomfortable when they are "away from home," something that happens very often in modern society, which is based on collaboration and "facilitates" (almost demands) mobility.

The concept of virtual residence (Beslay & Punie, 2002) describes the evolution of the physical home into the smart, digital home that will exist within an AmI space. It consists of the smart and connected home infrastructure and the objects therein; the online lives of people, families, households; and the services that support interaction, mobility, and interoperability between different AmI environments. Note that the concept uses most of the notions that apply to "physical" residence, such as borders, markers, activities, and so forth, and expands them to the digital space. Thus, in a virtual residence, one has to define digital borders to delimit the use of digital information, markers that describe allowable interactions with digital media, bridges (i.e., sensors and actuators) between the physical and digital spaces, and policies that ensure a balance between privacy and security, and enhance users' identities (Daskala & Maghiros, 2006).

Mobility can be supported by the virtual residence infrastructure. The aim is to make people feel "at home," even when they are physically located in a different place, by providing them access to content and services in such a way that, in the new location, they can either port or continue the tasks they usually perform at home with minimum disruption, or develop the feeling of being aware of or with the other home residents. These requirements extend beyond mere remote control of home services, which are currently supported via Internet-based applications.

Types of mobility that this concept can support include:

- **Migration of tasks:** The user can continue executing a task, even when he/she moves between AmI spheres. A simple example is the handover service offered by mobile phone operators. Within a rich AmI environment, this would require the restoration of the processes used by the task using the services available in every sphere.
- **Environment porting:** The user reconstructs his/her working, living, and so forth, environment within a new sphere using the objects in the sphere.
- **Mobility of people:** The user, being either on the move or away from the virtual residence, still has access to the services it offers (i.e., conceptually, the virtual residence borders "expand" to still contain the user).

ENGINEERING APPROACH

In order to build UbiComp systems from autonomous, resource-constrained, possibly heterogeneous components making optimum use of distributed intelligence embedded in the periphery, one must develop theories, technologies, and scientific communities that are undoubtedly interdisciplinary. These systems are modelled as self-aware and self-reconfigurable symbiotic ecologies where artificial beings and humans coexist. The applications include autonomous software running on autonomous devices. Social interactions arise among the different elements and adaptation to unforeseen (at design time) situations encountered in dynamic environments is needed.

Basic Building Blocks and Emergent Behaviour

When trying to define the basic building block, one is confronted with questions on i) which should be the basic building block, ii) what structural and functional properties it should encompass, iii) how it could interact with the others, and iv) how it could be realised. From a technology point of view, one would consider as basic building block every self-sustained digital (h/w or s/w) artefact with certain functionality that (a) can operate without the contribution of others, and (b) can interact with others. This definition includes robots with predefined specialised capabilities, but also sensors, motors, computational sources, and so forth. In all cases, as a result of interactions among basic building blocks, it is possible for emergent ecologies to be formed, which exhibit capabilities not found in the individuals.

One approach is to consider basic building blocks as hyperobjects (Mavrommati & Kameas, 2003) and treat everyday objects as communicating tangible components. This approach has been implemented in the gadgetware architectural style (GAS) (Kameas et al., 2003), which provides a platform (GAS-OS) that supports the composition of applications from interacting autonomous artefacts (called eGadgets) with the use of the plug-synapse metaphor (to be described later). This approach scales both "upwards" towards the assembly of more complex objects, and "downwards" towards the decomposition of eGadgets into smaller parts.

In dynamic environments, an individual must be reactive, that is, it must be responsive to events that occur in its environment, where these events affect either the individual's goals or the assumptions, which underpin the procedures that the individual is executing in order to achieve its goals. As a result, it is hard to build a system that achieves an effective balance between goal-directed and reactive behaviour (Wooldridge, 1999). Furthermore, when the construction of the individuals is based on the composition of primitive behaviours, the issues of how to select potentially the correct behaviours in different circumstances, and how to resolve conflicts between them, are raised.

The primitive behaviours approach considers that all the (individual) behaviours run in parallel and, depending on the stimuli of the environment, some of them manifest themselves by enabling a suppression mechanism and taking control of the actuators. However, this technique requires a predefined and exhaustively tested set of implicit rules (usually encoded into finite state automata) of firing priorities. Thus, this technique does not scale well even in a moderate number of primitive behaviours, and it lacks learning even in very often tasks.

In order to apply the well-established primitive behaviours approach in swarm societies that can learn and evolve, component-oriented principles and practices could be employed. Synthetic behaviour control mechanisms could be developed based on bio-inspired approaches like spiking neural networks. These behaviour control mechanisms responsible for the arbitration and/or the composition of the primitive behaviours could also be the subject of learning and evolution. The individuals may exhibit varying behaviour: perceiving/exploring their environment, selectively focusing attention, initiating and completing several tasks. The learning and evolution could be studied and investigated at both the individual and social levels. In this case, the focal point must be the components of behaviour control mechanisms. The outcome could contribute to a novel dynamic and adaptive architecture of swarm systems that exploits the global effects through local rules/behaviour.

AmI Spheres and Collective Behaviour

The realisation of an AmI sphere and the supporting computational environment raises several issues fundamentally pertaining to sphere architectural design and real-time perception-deliberation-action loop. The key features of an AmI sphere are its omnipresence via emergence, polymorphism, and adaptation.

In building a swarm system, communication plays a pivotal role. A flexible and lightweight approach is the indirect (stigmergic) communication. The essence of stigmergy is that the individual modifies a local property of the environment that, subject to environmental physics, should persist long enough to affect the individual's behaviour later in time. It is the temporal aspect of this phenomenon that is crucial for emergent collective behaviour (collaborative exploration, building and maintenance of complex insect nest architectures, etc.) in societies of ants, agents, and robotics. Thus, the individuals could be provided with the proper periphery (actuators/sensors), enabling them to emit/perceive electromagnetic signals and emulate the biological "quorum sense" signals. Such a quorum sense communication may be based on an application-specific vocabulary that will be encoded in the signal. The specifics of the temporal modulation aspect of this "quorum sense signal" will come from theoretical biology and existing results. Additionally, the frequency of the signal will be determined after studying the combined influence of the physical medium properties, the range and interference constraints, power requirements, and the size of available hardware components.

Awareness and Presence

Stigmergic signals are used to denote presence. In Sheridan, (1992) there are proposed three categories of determinants of presence: (1) the extent of sensory information presented to the participant, (2) the level of control the participant has over the various sensor mechanisms, and (3) the participant's ability to modify the environment. In Lombard and Ditton, (1997) presence is defined as the "perceptual illusion of nonmediation", that is, the extent to which a person fails to perceive or acknowledge the existence of a medium during a technologically mediated experience. One can roughly identify two types of presence: physical and social. *Physical presence* refers to the sense of being physically located in mediated space, whereas *social presence* refers to the feeling of being together, of social interaction with a virtual or remotely located communication partner.

Presence is used in our approach as a universal concept that applies to all actors (i.e., people, agents, objects) within a sphere, although each actor will recognize different types of signals, use different mediums, and employ different mechanisms to perceive presence. For example, digital entities will exchange stigmergic signals using digital traces, while people will prefer visual or auditory cues. When people to object mediated interaction must be supported, the "valid" stigmergic signals will be constrained by the sensors that the object embeds. To achieve object to people stigmergic communication, we adopt principles from awareness systems and social intelligence. Awareness systems are a class of computer-mediated communication (CMC) systems that support individuals to maintain, with low effort, a peripheral awareness of each other's activities. The Casablanca project (Hindus et al., 2001) and the ASTRA project (Markopoulos, Romero, van Baren, Ijsselsteijn, de Ruyter, & Farschian, 2004) are examples of early awareness systems for the home.

INTERACTING WITH AMI SPHERES

The AmI spheres constitute a dynamic distributed system composed of artefacts with finite sets of capabilities (services) offered usually through proprietary user interfaces. People interact with an AmI sphere in two levels:

- The task level, whereby they will have to use each individual artefact in order to make use of the collective AmI sphere capabilities
- The metatask level, whereby they will have to compose, decompose, or otherwise edit AmI spheres

When interacting with an AmI sphere, people are in fact using the artefacts that compose it (i.e., they are simply acting, not interacting). This is as close as we can get to the notion of calm technology promoted by M. Weiser, who stated that the most profound technologies are those that disappear in the background (Weiser, 1993). This view is directly inspired from Heidegger's theory of "dasein," which states that people are thrown in the world and are always engaged with acting within it to accomplish their tasks. In this view, technological tools disappear in the background in favour of tasks-at-hand; tools only appear when the task accomplishment procedure breaks down, that is, when something goes wrong.

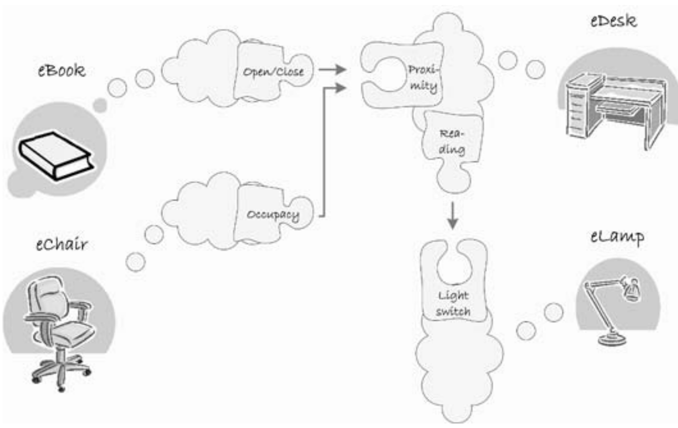
Artefacts have to demonstrate their affordances, both in the physical environment (for people to be able to use them) and to the digital space (so that other artefacts, agents, and so forth, will be able to interact with them). Then, the state of each artefact must be made visible/available for the same reasons (although the procedure used to compute the state should be internal to the artefact). AmI spheres introduce two factors that cause the breakdown of the existing task models: people will have to make sure that they can still carry on with ordinary tasks, and they will have to become familiar with the new affordances of the artefacts. In addition to adapting their skills for using artefacts, people will have to develop skills for using the computing properties of their new environments as well (Mavrommati & Kameas, 2004).

The GAS Approach

To deal with this, GAS adopts a layered architecture that transparently supports composing and using AmI spheres (called eGadgetWorlds) from autonomous artefacts (called eGadgets), which can be objects, services, or both (Kameas et al., 2003). To enable composition of AmI spheres, GAS proposes the plug-synapse model: a "plug" is the manifestation of a property, capability, or service in a semantically-rich way, and a "synapse" is a communication established between compatible plugs. For example, a TV set may offer the display service and a digital camera may establish a synapse in order to output images; a chair may offer the capability to recognize whether a person is seated on it and a table lamp may use it to switch itself on, and so forth. Clearly, this concept scales well, as more complex "plugs" can be defined as compositions of simpler "plugs," either at an artefact or sphere level. For example, a "reading plug" for an office AmI sphere (Figure 2) may be defined by combining specific plugs from a chair, a table, a lamp, and a book (in fact, we refer to their artefact counterparts) in such a way that when someone is seated on the chair and the chair is located close enough to the table and a book is opened on the table, then the lamp is switched on. With the help of the room (considered as an artefact), the system could also recognize who is seated on the chair and switch on automatically his/her reading profile; then, the "reading plug" could be used by the room to redirect phone calls so as not to disturb the user unless necessary.

An interesting case appears when the AmI sphere breaks down. Consider, for example, the case where the desk lamp is broken. Then the system can either inform the user and wait for his/her action, or search for a similar service in the environment (for example, the sphere, with the help of the AmS system, can locate the room lamp and switch it on). In the latter case, it is necessary that all artefacts hold an internal description of their services and goals; and that these descriptions are compatible. GAS includes a multilayered ontology that describes artefact "plugs" and rules of usage (i.e., constraints) using a commonly available core ontology of basic terms. The use of ontology makes possible the communication between heterogeneous eGadgets, and helps in achieving a shared understanding (as described in Habermas, 1984). Emergent behaviour of this type is a direct result of the ability of eGadgets to communicate

Figure 2. An example eGadgetWorld that implements the "reading office" sphere



in socially meaningful ways, as described in their hierarchy of basic behaviours.

This definition supports emerging functionality because (a) artefacts are self-sufficient and their plugs are described in a functionally independent way, (b) not all synapses need to be known from the start, (c) new synapses may be added or existing synapses may be deleted, for example, as an artefact may move outside the sphere (that is, outside the range of the wireless network), and (d) experience may be recorded in local artefact ontologies and appear in the form of higher level plugs (the use of a common core ontology available in the sphere's environment ensures the compatibility of plug definitions).

Plugs and synapses are managed by GAS-OS, a distributed middleware platform that takes care of resource management and communication (Figure 3). Thus, an AmI sphere is defined as a GAS-OS application; all eGadgets in it run GAS-OS; compatible plugs of these eGadgets are engaged in synapses to provide collective sphere functionality.

GAS regards each artefact or mobile device as an autonomous component of an AmI sphere. Although GAS-OS has to run on each artefact or mobile device to ensure compatibility, each artefact can locally and transparently manage its resources. Other approaches support the downloading of software representatives of artefacts, either close (Siegemund & Krauer, 2004) or remote (Lopez de Ipina *et al*., 2005), into a mobile device, thus making the device to assume the role of a superartefact capable of running a component framework. GAS aims to maintain functional autonomy of artefacts; moreover, mobile devices are also considered as artefacts from an interaction perspective.

Supporting Tasks and Metatasks

Within an AmI sphere composed as an eGadgetWorld, a user may perform his/her tasks simply by using the artefacts or mobile devices therein. We do not propose to embed screen-based interfaces on every artefact, or to use a computer as a sphere master, as these would greatly alter the affordances of the artefacts and consequently, have a negative effect on people's capacity to form new task models. Another unwanted consequence is that these artefacts would no longer be functionally autonomous.

The issues that pertain to the individual artefact user interfaces and the ways their affordances can be manifested will not concern us here; a treatise of these can be found in Mavrommati and Kameas (2003). Instead, we are concerned with the collective sphere interface. To this end, we adopt the basic notions and goals of UI plasticity (Calvary, Coutaz, & Thevenin, 2002), although we are not

Figure 3. GAS-OS layered design

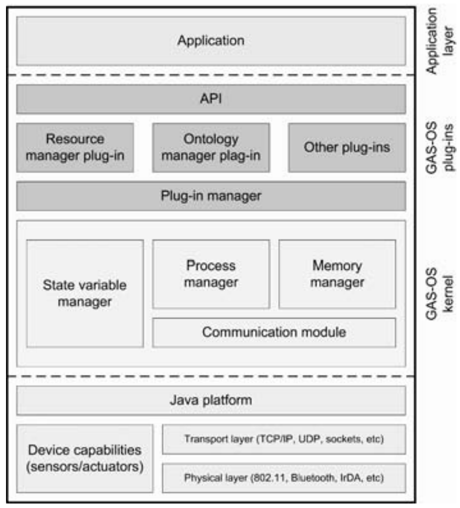


Figure 4. AmI sphere configuration tools

concerned about UI migration and distribution, as these apply to centralized UIs. In the case of AmI spheres, the issues are to conduct a coherent dialogue composed of user actions within the sphere and eGadgets' responses, and to preserve a distributed but meaningful dialogue state. To achieve this, each eGadget must be aware of the state of other eGadgets in the sphere. This can be achieved by exchanging information through the synapses that compose the sphere. By processing the combined perception of the states of itself, peer artefacts (connected via synapses), and the environment (through stigmergic/awareness signals), and applying its architecture of subsumed behaviours, each artefact is able to locally maintain a dialogue state that is compatible with the AmI sphere dialogue state.

GAS offers a set of tools that support the metatasks of creating and editing spheres (Figure 4). These editors run on mobile devices (i.e., PDAs and laptops), and have been positively evaluated by nonexpert users (Mavrommati, Kameas, & Markopoulos, 2004).

The Subsumption Architecture

The proposed conceptual framework extends the GAS approach by allowing the eGadgetWorld management tasks to be dynamically performed by the AmS (though the direct user involvement is not prohibited) based on the observed user habits, the available artefacts and services, and the social rules in context. The issue that arises with this approach is the uniform support for individual functionality and "social interactions" between the sphere members (artefacts, agents, and people).

Our modelling approach achieves uniformity and coherence because it uses the notion of basic behaviours to represent both functional and social capacities of eGadgets. Thus, the decision-making module of the middleware of each eGadget is composed of two kinds of basic behaviour modules: those that implement its core functionality, and those that realize the context of social intelligence. The former use the data gathered by the eGadget sensors to calculate the object's state and to decide a set of (re)actions. The latter use sensor data, as well as synapse input, to determine context of operation and to select the most appropriate action in the list. All basic behaviours are placed in the same hierarchy, with social intelligence behaviours having greater precedence over functional behaviours; thus, allowing the eGadget to realise the most socially intelligent response. Using this two-level selection mechanism, we ensure independence in the determination of local state and response, while we achieve a socially driven eGadget behaviour.

Examples of functional basic behaviours are "turn light on," "produce specific sound," "move towards a specific direction," and so forth; these depend on the actuators of the eGadget and determine the affordances the eGadget offers. This set also contains the basic behaviours "form synapse" and "learn", which ensure that affordances such as composeability and changeability are supported. To deal with possible heterogeneity in signal definition, each eGadget uses a local ontology to translate incoming signals through synapses.

Regarding social behaviour, we consider basic social behaviours drawn from the social intelligence studies,for example, benevolence, nonintrusion, altruism, responsibility, antagonism, empathy, emergency, and so forth. The subsumption scheme contains, in a hierarchical structure, several degrees of sociality, from nonsociality at all in the lower level to high sociality in the top level, as well as the suppression between the social behaviours. The mechanism that implements the subsumption scheme and is responsible for the arbitration of the social behaviour enablement, according to the context it has to deal with, resides in the abstract level of the AmS. Thus, the sociality of the basic building blocks participating in an emergent society is defined without requiring representation or reasoning capabilities from the artefacts.

The GAS Symbiotic Interaction Metaphor

In order to describe our proposed metaphor for interacting with UbiComp applications composed from communicating artefacts, let us first make explicit some basic assumptions:

- User inhabits an AmI space that contains artefacts, having a physical presence and offering digital services
- User forms a plan to achieve a goal he/she has in mind; in this stage, the plan probably consists of steps and subgoals, some of which may not be conscious at all
- User tries to realise his/her plan by combining services offered by the artefacts in his/her environment; in this endeavour, user can only be aware of the affordances of the artefacts and tries to use them accordingly

First of all, GAS supports the following metatasks, using the GAS editor:

- User can query the services and capabilities of each eGadget
- User composes an eGadgetWorld by combining eGadgets capabilities using the plug/synapse model

- User can have an overview of the existing eGadgetWorlds and even edit or delete any of them

Then, we propose the following interaction metaphor:

- The state of each eGadget is communicated using its actuators
- User simply uses each eGadget based on its affordances, directly affecting its local state
- As a consequence, the eGadget communicates the new state to the eGadgets it is connected to via its synapses
- Peer eGadgets calculate new local states (thus user indirectly affects them) and communicates them using their actuators; each eGadget decides the form of communication using its local subsumption architecture, its local state, and the context it perceives via its local sensors and peer eGadgets states (via its synapses)
- A new global eGadgetWorld state emerges as a consequence of local state changes of all the eGadgets in the eGadgetWorld
- The new global state is communicated to the user by all eGadgets in the eGadgetWorld

In this approach, we must make a few remarks. Firstly, because this is a symbiotic ecology, there is no centralized decision-making component. All eGadgets are considered as peers, and each one is responsible for local decision making and acting by taking into account local and global information (here "global" is restricted to those eGadgets that have synapses with it). However, the eGadgets that compose an eGadgetWorld, if they have this basic behaviour, may choose to elect representatives, that is, eGadgets that will act as eGadgetWorld "leaders." This process is supported by contemporary communication protocols (i.e., Wi-Fi) and ensures that the eGadgetWorld will remain functional, even if some secondary or weak eGadgets are not occasionally operating.

Secondly, because the local eGadget state is communicated to other peer eGadgets via the synapses, and triggers changes in their local states, which are also communicated to peer eGadgets, there exists the risk of the eGadgetWorld falling in an infinite loop of recursive global state changes. This falls within the scope of our modelling approach and can be avoided using two measures. Firstly, synapses are directed: if a synapse exists from eGadget A to eGadget B, this means that only changes in the state of eGadget A are communicated to eGadget B, and not the other way round. In addition, when composing or editing an eGadgetWorld, the GAS tools offer the user the ability to send a "ping" message that propagates to all eGadgets in the eGadgetWorld, thus making clear if any loops exist.

AN EXAMPLE SCENARIO

Sonia is a 36-year old single mother who lives in Athens with here two children. She is a hard-working expert employee who is overcommitted with her children. These days she is working on a report, and today is the deadline for its submission due to a company meeting at 20:00 local time in Brussels. However, she could not travel far from Athens and thus, she could not participate in the meeting. Instead, her colleague, Steve, will present her report. Undoubtedly, she is anxious about the results. As the meeting will take long, she cannot stay awake, but she needs to hear the news from Steve as soon as she wakes up in the morning.

It is 16:30 in Athens (+1:00 CET) and Sonia works in her office. She makes the last minute changes in her report, but she realises that she forgot some important handwritten notes in her home. Additionally, she must take her children to their ballet class. She must urgently leave her office, but also she must finalise and submit the report. Fortunately, she has a little more time until 21:00 local time.

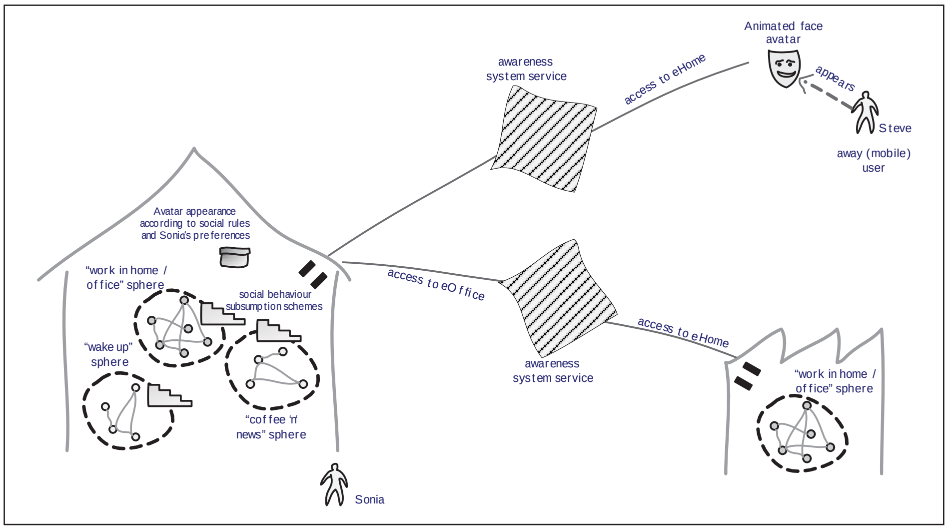
Additionally, Sonia has "an ace in her sleeve"; she likes technology, and has created several eGadgetWorld as AmI spheres to make her overcommitted life easier (Figure 5):

- Her "work-in-home/office" sphere includes the eBook, eNotebook, eChair, eDesk, eLamp, and ePC eGadgets. The functionality of this sphere is similar to the one described in "The GAS Approach" section with the extension that the ePC eGadget provides a synchronisation service concerning a shared repository between different and remote PCs (e.g., home PC and office PC).
- The "wake up" sphere consists of the following eGadgets: an alarm clock, a bed mattress, the window blinds, and a room light. The eGadgets in this sphere have plugs through which they offer access to their properties and services. For example, the clock is equipped with a light sensor whose reading is made available through the plug "luminosity"; the bed mattress has weight and pressure sensors and can decide whether there is someone "lying upon" or not; the window blinds offer the plug "open," which can be used to lift them to a specified height; finally, the room light offers the plug "on-off." This eGadgetWorld, clearly a ubiquitous computing application, is set up to gradually increase the amount of light in the room (until Sonia gets up from bed) when it is time for her to wake up, first by opening the window blinds to let natural light come in, and, if it is still dark (i.e., consider a winter's day when the lighting remains lower than a certain threshold), switch on the room light. All Sonia had to do is to synapse the "luminosity" plug of the alarm clock with "open" plug of the blinds and the "on-off" plug of the room light, synapse the "alarm on-off" plug with the "person lying" plug of the mattress (so that when she stands up from the bed, the alarm is switched off and the blinds stop opening), and finally set the alarm clock.
- In a third, simpler eGadgetWorld, she has already synapsed the mattress with her slippers, and the latter with the coffee maker, thus starting coffee brewing when she gets up and steps into the slippers. Now, she needs to extend it, with a few more synapses between her bathroom mirror and Steve's avatar. She names this eGadgetWorld "coffee 'n' news." It is worth mentioning that Sonia uses this mirror in another eGadgetWorld as a display to read the morning news and the weather as she gets ready for work.

It is 18:00 local time and Sonia prepares dinner. She has already put the children in their room to play. She exploits the time until the dinner is ready and moves to her desk in the penthouse. The "work in home/office" sphere is active and when she sits on her eChair, the ePC pops up a sign to inform her of an urgent task to be completed. In fact, the awareness service between the home and office spheres triggered the synchronisation synapse, and the home PC informed about the task by the office PC and also downloaded the unfinished report into the shared repository. Sonia takes a look at her handwritten notes, completes the report, and sends it to Steve; the home PC closes the task, and the awareness service will inform the office PC about it when Sonia goes to her office and starts using her PC. Now, it is time for dinner. A couple of hours later, the children are in their beds and Sonia feels tired and sleepy.

The "wake up" sphere is active and the next morning, at 06:45 local time, as the sun starts rising, the light sensors are triggered and the luminosity plugs arrange the bedroom illumination. As Sonia gets up and steps into the slippers, the coffee starts brewing, as the "coffee 'n' news" sphere is active. Now Sonia is in her bathroom in front of the mirror and brushes her teeth. The "news" synapse with the eMirror, which she made yesterday, performs a connection with Steve's avatar; he is already awake as he must catch the morning flight to Athens and also cannot wait to inform Sonia about the yesterday's meeting. Clearly, he did not forget to synapse his avatar with Sonia's eHome sphere. Although the eMirror is able to provide live video services, it does not do so this time; it is not socially proper to "intrude" on Sonia's private spaces. The social behaviour service, endowed with the AmS, informs the avatar about the pri-

Figure 5. AmI spheres and interactions



vacy restrictions and allows it to appear only as a "talking box." Now, Sonia gets the good news about the Brussels meeting. She can enjoy her morning coffee. A good day just started.

CONCLUSION

As everyday objects are being enhanced with sensing, processing, and communication abilities, the near future of our everyday living/working is indicated by a high degree of complexity. The emergent complexity concerns the machine-machine and human-machine interactions as well as the provided services aimed at end users and at other machines. Into this rapidly changing AmI environment, new requirements and research issues arise, and the need for a conceptual analysis framework is apparent. This work attempts to introduce a bio-inspired world model that draws features from natural systems, and applies them into symbiotic ecologies inhabited by both humans and artefacts. Furthermore, it introduces a high-level framework of AmI spaces that encloses the fundamental elements of bio-inspired self-aware emergent symbiotic ecologies.

REFERENCES

Baker, C., R., Markovsky, Y., van Greunen, J., Rabaey, J., Wawrzynek, J., & Wolisz, A. (2006 July). ZUMA: A platform for smart-home environments. In *Proceedings of the 2nd IET Conference on Intelligent Environments.* Athens, Greece,

Beslay, L., & Punie, Y. (2002). The virtual residence: Identity, privacy and security. *The IPTS Report, Special Issue on Identity and Privacy, 67*, 17-23.

Bonabeau, E., Dorigo, M., & Theraulaz, G. (1999). *Swarm intelligence: From natural to artificial systems.* Oxford University Press.

Brooks, R. A. (1991). Intelligence without representation. *Artificial Intelligence*, *47*, 139–159.

Brooks, R. A. (2002). *Robot: The future of flesh and machines.* London: Penguin Books.

Bullock, S., & Cliff, D. (2004). *Complexity and emergent behaviour in ICT systems* (Tech. Rep. No HP-2004-187). Semantic & Adaptive Systems, Hewlett-Packard Labs.

Calvary, G., Coutaz, J., & Thevenin, D., (2002). A unifying reference framework for the development of plastic user interfaces. In *Proceedings of EHCI01* (pp. 173-192). Toronto, LNCS 2254.

Dahlgren, T., Epperly, T., & Kumfert, G. (2003). *Babel users' guide*, version 0.8.4. CASC Technical Publication, Lawrence Livermore National Laboratory.

Daskala, B., & Maghiros, I. (2006 July). Digital territories. In *Proceedings of the 2nd IET Conference on Intelligent Environments.* Athens, Greece.

Drossos, N., Mavrommati, I., & Kameas, A. (2007). Towards ubiquitous computing applications composed from functionally autonomous hybrid artefacts. In N. Streitz, A. Kameas, & I. Mavrommati (Eds.), *The disappearing computer book*. Springer-Verlag, (to appear in 2007).

Edwards, W., K., Newman, M., W., Sedivy, J., Smith, T., & Izadi, S. (2002 September). Challenge: recombinant computing and the speakeasy approach. In *Proceedings of the Eighth Annual International Conference on Mobile Computing and Networking* (MobiCom 2002) (pp. 279-286). ACM Press, New York.

Foster, I., Jennings, N., & Kesselman, C. (2004 July). Brain meets brawn: Why grid and agents need each other. In *Proceedings of the 3rd International Joint Conference on Autonomous Agents and Multi-Agent Systems, (AAMAS 04).* New York.

Habermas, J. (1984). *The theory of communicative action (Vol. 1-2)* (translator T. McCarthy). Boston: Beacon Press.

Hindus, D., Mainwarin, S. D., Leduc, N., Hagstro A. E., & Bayley, O. (2001, March). Casablanca: Designing social communication devices for the home. In *Proceedings of the SIG-CHI on Human factors in computing systems (CHI 2001)* (pp 325–332). Seattle, WA.

Humble, J., Crabtree, A., Hemmings, T., Akesson, K. P., Koleva, B,. Rodden, T., & Hansson, P. (2003 October). Playing with the bits: User-configuration of ubiquitous domestic environments. In *Proceedings of UBICOMP 2003* (pp. 256-263). Springer-Verlag.

IST Advisor Group. (2001). *Scenarios for ambient intelligence in 2010*. European Commission Community Research. Retrieved on September 12, 2006, from http://cordis.europa.eu/ist/istag-reports.htm

Kameas, A., Bellis, S., Mavrommati, I., Delanay, K., Colley, M., & Pounds-Cornish, A. (2003 March). An architecture that treats everyday objects as communicating tangible components. In *Proceedings of the First IEEE International Conference on Pervasive Computing and Communications*, (PERCOM2003). Texas-Fort Worth.

Kameas, A., Mavrommati I., & Markopoulos, P. (2004). Computing in tangible: Using artifacts as components of ambient intelligent environments. In G. Riva, F. Vatalaro, F. Davide, & M. Alcaniz (Eds.), *Ambient intelligence: The evolution of technology, communication and cognition* (pp. 121-142). IOS Press.

Kennedy, J., & Eberhart, R. (2001). *Swarm intelligence*. Morgan Kaufmann Publishers.

Knoll, A., & de Kamps, M., (Eds). (2004). *Roadmap of neuro-IT development.* EU Neuro-IT Network of Excellence. Retrieved on September 12, 2006, from http://www.neuro-it.net/NeuroIT/Roadmap

Larson, W., Norris, B., Ong, E. T., Bernholdt, D. E., Drake, J. B., El Wasif, W. R., Ham, M. W., Rasmussen, C. E., Kumfert, G., Katz, D. S., Zhou, S., DeLuca, C., & Collins, N. S. (2004). Components, the common component architecture, and the climate/ocean/weather community. In *Proceedings of the 20th International Conference on Interactive Information and Processing Systems (IIPS) for Meteorology, Oceanography, and Hydrology*, 84th American Meteorological Society Annual Meeting.

Lindwer, M., Marculescu, D., Basten, T., Zimmermann, R., Marculescu, R., Jung, S., & Cantatore, E. (2003 March). Ambient intelligence visions and achievements: Linking abstract ideas to real-world concepts. In *Proceedings Design Automation & Test in Europe (DATE).* Munich, Germany.

Lombard, M., & Ditton, T. (1997). At the heart of it all: The concept of presence. *Journal of Computer Mediated-Communication [On-line], 3.* Retrieved on January 17, 2007, from http://www.ascusc.org/jcmc/vol3/issue2/lombard.html

Lopez de Ipina, D., Vaszquez, I., Garcia, D., Fernandez, J., & Garcia, I. (2005 October). A reflective middleware for controlling smart objects from mobile devices. In *Proceedings of the 3rd European joint Symposium on Systems on Chip and Ambient Intelligence (sOc-EUSAI)* (pp. 213-218). Grenoble, France.

Markopoulos, P., Romero, N., van Baren, J., Ijsselsteijn, W., de Ruyter, B., & Farschian, B. (2004 April). Keeping in touch with the family: Home and away with the ASTRA awareness system. In *Proceedings of the International Conference for Human-Computer Interaction, (CHI2004)* (pp. 1351-1354), Vienna, Austria.

Mataric, M., J. (1995). Designing and understanding adaptive group behaviour. *Adaptive Behaviour, 4*(1):51-80.

Mavrommati, I. & Kameas, A. (2003). The evolution of objects into hyper-objects. *Personal and Ubiquitous Computing 779, ACM, 7*(3-4), 176-181.

Mavrommati, I., & Kameas, A. (2004 April). The concepts of an end user enabling architecture for ubiquitous computing. In *Proceedings of the 2nd International Conference on Pervasive Computing (Pervasive 2004).* Vienna.

Mavrommati I., Kameas A., & Markopoulos P. (2004 May). An editing tool that manages device associations in an in-home environment. In *Proceedings of the 2nd International Conference on Appliance Design (2AD)* (pp. 104-111), HP Labs, Bristol, UK, 11-13 May 2004.

Mitchell Waldrop, M. (2003). *Autonomic computing—The technology of self-management.* Retrieved on September 12, 2006, from http://www.thefutureofcomputing.org

Mozer, M. C. (2005). Lessons from an adaptive house. In D. Cook R. & Das, R., (Eds.), *Smart environments: Technologies, protocols, and applications* (pp. 273-294). Hoboken, NJ: J. Wiley & Sons.

Norman, D. A. (1998). *The invisible compute.* MIT Press.

Russell, S., & Norvig, P. (2003). *Artificial intelligence: A modern approach* (2nd ed.). Prentice Hall.

Sheridan, T. B. (1992). Musings on telepresence and virtual presence. *Presence: Teleoperators and Virtual Environments, 1*, 120-125.

Siegemund, F., & Krauer, T. (2004 November). Integrating handhelds into environments of cooperating smart everyday objects. In *Proceedings of the 2nd European Symposium on Ambient Intelligence (EUSAI 2004)* (pp.160-171). Eindhoven, The Netherlands: Springer-Verlag.

Thorndike, E. L. (1920). Intelligence and its uses. *Harper's Magazine*, *140*, 227-235.

Truong, K. N., Huang, E. M., & Abowd, G. D. (2004 September). CAMP: A magnetic poetry interface for end-user programming of capture applications for the home. In *Proceedings of the Sixth International Conference on Ubiquitous Computing, (Ubicomp 2004)* (pp. 143-160). Nottingham, England.

Vernon, P. E. (1933). Some characteristics of the good judge of personality, *Journal of Social Psychology*, *4*, 42–57.

Watkins, C. J. C. H., & Dayan, P. (1992). Q learning. *Machine Learning*, *8*,279–292.

Weiser, M. (1993). Some computer science issues in ubiquitous computing. *Communications of the ACM*, *36*(7),75–84.

Wooldridge, M. (1999). Intelligent agents. In G. Weiss (Ed.), *Multiagent systems: modern approach to distributed artificial intelligence* (pp. 27-77). MIT Press.

KEY TERMS

Ambient Intelligence: A set of [emergent] properties of an environment that we are in the process of creating; it is more an imagined concept than a set of specified requirements (IST Advisor Group, "Ambient intelligence: From vision to reality." Retrieved on September 16, 2006, from http://cordis.europa.eu/ist/istag-reports.htm). In particular, AmI puts the emphasis on user friendliness, efficient and distributed services support, user empowerment, and support for human interactions. This vision assumes a shift away from PCs to a variety of devices that are unobtrusively embedded in our environment, and that are accessed via intelligent interfaces (Retrieved on September 16, 2006, from http://en.wikipedia.org/wiki/Ambient_intelligence).

Ambient Sphere: Ecology of artificial entities coexisting unobtrusively with humans and performing collaborative tasks through a continuous evolvable process concerning both their physical and social cognitive growth.

Ambient System: A distributed platform that supports the instantiation of ambient spheres.

Basic Behaviours: Predefined functions of the artefacts that are either enabled as reactions to external events or are continuously pursued to be fulfilled.

Basic Building Block: Modelling abstraction representing self-sustained entities that are members of an AmI sphere.

Hyper Objects (eGadgets): Ordinary objects that are commonly used for everyday, even mundane tasks (objects such as tables, chairs, cups, shelves, lights, carpets, etc..), and which in the future, can be enhanced with communication, processing, and sometimes sensing abilities (Mavrommati & Kameas, 2003).

Plug-Synapse Model: A conceptual abstraction that enables uniform access to eGadget services/capabilities/properties and allows users to compose applications that realize a collective behaviour in a high-level programming manner (Mavrommati & Kameas, 2004)

Social Intelligence: "The ability to understand and manage men and women, boys and girls—to act wisely in human relations." (Thorndike, 1920). According to a broader definition, social intelligence is "… a person's ability to get along with people in general, social technique or ease in society, knowledge of social matters, susceptibility to stimuli from other members of a group, as well as insight into the temporary moods of underlying personality traits of strangers" (Vernon, 1933).

Swarm Intelligence: "… an alternative way of designing intelligent systems, in which autonomy, emergence and distributed functioning replace control, preprogramming, and centralisation" (Bonabeau et al., 1999).

Ubiquitous Computing: "…the method of enhancing computer use by making many computers available throughout the physical environment, but making them effectively invisible to the user" (Weiser, 1993).

Section II
Novel Interaction Techniques for Mobile Technologies

This section focuses on the innovative possibilities for interaction with mobile technologies. Starting with a potential classification scheme for mobile interaction techniques, this section looks at a number of novel interaction techniques such as text entry, speech-based input, and audio and haptic interaction for mobile devices. Chapters are included that introduce the concept of unobtrusive interaction and the use of EMG signals to achieve subtle interaction. This section concludes with a look at visual means of interaction, from camera-based input, through 3-D visualisation and the presentation of large data sets using starfield displays, to projected displays for collaborative interaction.

Chapter XXIV
The Design Space of Ubiquitous Mobile Input

Rafael Ballagas
RWTH Aachen University, Germany

Michael Rohs
Deutsche Telekom Laboratories, Germany

Jennifer G. Sheridan
BigDog Interactive Ltd., UK

Jan Borchers
RWTH Aachen University, Germany

ABSTRACT

The mobile phone is the first truly pervasive computer. In addition to its core communications functionality, it is increasingly used for interaction with the physical world. This chapter examines the design space of input techniques using established desktop taxonomies and design spaces to provide an in-depth discussion of existing interaction techniques. A new five-part spatial classification is proposed for ubiquitous mobile phone interaction tasks discussed in our survey. It includes supported subtasks (position, orient, and selection), dimensionality, relative vs. absolute movement, interaction style (direct vs. indirect), and feedback from the environment (continuous vs. discrete). Key design considerations are identified for deploying these interaction techniques in real-world applications. Our analysis aims to inspire and inform the design of future smart phone interaction techniques.

INTRODUCTION

Today, mobile phones are used not just to keep in touch with others, but also to manage everyday tasks, to share files, and to create personal content. Consequently, our mobile phones are always at hand. Just as Mark Weiser suggested in his vision of ubiquitous computing, the ubiquitous nature of mobile phones certainly does make them "*blend into the fabric of our everyday lives*" (Weiser, 1991).

Technology trends show an increasing number of features packed into this small, convenient form factor. Smart phones already have eyes (camera), ears (microphone), and sensors to perceive their environment. However, their real power, as Weiser pointed out, comes not just from one device, but from the interaction of all of them. Our interest is in showing how modern mobile phones, which resemble Weiser's "*tabs*," can be used as interaction devices for our environment. Within this environment, emphasis will be placed on interactions with public and situated displays (O'Hara, Perry, & Churchill, 2004) – what Weiser called "*boards*."

The range of input and output (I/O) capabilities for modern mobile phones is broad. Keypad, joystick, microphone, display, touch-screen, loudspeaker, short-range wireless connectivity over Bluetooth, WiFi, or infrared, and long-range wireless connectivity via GSM/GPRS and UMTS all provide multiple ways of interacting with our phones. These multiple I/O capabilities have increased our ability to use mobile phones to control resources available in our environment, such as public displays, vending machines, and home appliances.

Could this ubiquity mean that mobile phones have become the default input device for ubiquitous computing applications? If so, then mobile phones are positioned to create new interaction paradigms, similar to the way the mouse and keyboard on desktop systems enabled the WIMP (windows, icons, menus, pointers) paradigm of the graphical user interface to emerge and dominate the world of personal desktop computing. However, before this potential is realized, first we must consider which input techniques are intuitive, efficient, and enjoyable for users and applications in the ubiquitous computing domain.

EXAMINING THE DESIGN SPACE OF INPUT DEVICES

Recent research demonstrates a broad array of mobile phone input techniques for ubiquitous computing application scenarios. To make sense of the cumulative knowledge, we systematically organize the input techniques to give insights into the design space. The design space is an important tool for helping designers of ubiquitous computing applications to identify the relationships between input techniques, and to select the most appropriate input technique for their interaction scenarios. Design spaces can also be used to identify gaps in the current body of knowledge and suggest new designs (Zwicky, 1967).

Looking to Foley, Wallace, and Chan's classic paper (Foley et al., 1984), we find a taxonomy of desktop input devices that are structured around the graphics subtasks that they are capable of performing (Position, Orient, Select, Path, Quantify, and Text Entry). These subtasks are the elementary operators that are combined to perform higher-level interface tasks, and will be elaborated upon in later sections. In this chapter, we structure our analysis of smart phones as ubiquitous input devices using this taxonomy. This analysis builds on classic design spaces (Buxton, 1983; Card, Mackinlay, & Robertson, 1991) and extends our own previous work (Ballagas, Ringel, Stone, & Borchers, 2003, Ballagas, Rohs, M., Sheridan, J., and Borchers, 2006) on the design space of input techniques. In our analysis, we blur the line between smart phones and personal digital assistants (PDAs) because their feature sets continue to converge.

Although Foley et al.'s analysis was completed with the desktop computing paradigm in mind, the subtasks in their analysis are still applicable to ubiquitous computing today. They naturally apply to situated display interactions; however, their applicability is not limited to graphical interactions.

In the following sections, each of Foley et al.'s subtasks will be examined in the context of mobile phone interactions. Foley et al.'s taxonomy uses the following input characteristics to further classify input techniques:

- **Feedback:** Continuous interactions describe a closed-loop feedback, where the user continuously gets informed of the interaction progress as the subtask is being performed. For example, when using a mouse, the current cursor position is continually fed back to the user. Discrete interactions describe an open-loop feedback, where the user is only informed of the interaction progress after the subtask is complete. For example, when selecting an object on a touch panel, the progress of the selection is not displayed until after the finger meets the surface to complete the selection of the desired item.
- **Interaction Style:** In direct interactions, input actions are physically coupled with the user-perceivable entity being manipulated (such as an image on a display). Physical coupling can be achieved when the feedback spatially coincides with the input action, or can be achieved at a distance if the user is manipulating a 3-D ray (such as with a laser pointer) that intersects directly with the entity being manipulated. To the user, this appears as if there is no mediation, translation, or adaptation between input and output.

In indirect interactions, user activity and feedback occur in disjoint spaces (e.g., using a mouse to control an on-screen cursor). Scaling and abstraction between input actions and feedback are often necessary in indirect interactions.

Position

During a POSITION task, the user specifies a position in application coordinates, often as part of a command to place an entity at a particular position. Positioning techniques can either be *continuous*, where the object position is continually fed back to the user, or *discrete*, where the position is changed at the end of the positioning task. Positioning tasks can further be differentiated using the directness of the interaction. In *direct* interactions, input actions are physically coupled with the object being positioned; in *indirect* interactions, user activity and feedback occur in disjoint spaces. We note that position could refer to screen position, or physical position in the real world. For example, the height of motorized window blinds can be adjusted using the position subtask.

The mobile phone has been used for positioning tasks in a variety of ways:

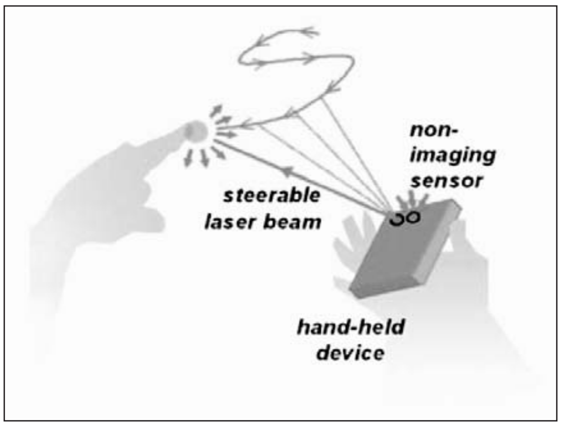
Continuous Indirect Interactions

1. **Trackpad:** A trackpad is a touch-sensitive surface that is used as a relative pointing device, standard in modern laptops. Remote Commander (Myers, Stiel, & Gargiulo, 1998) enables individuals to use the touch screen on a PDA as a trackpad to control the relative position of a cursor on a remotely situated display. In this interaction, the user's attention is concentrated on the situated display and no application-level feedback is provided on the PDA; thus, the functionality of the PDA is essentially reduced to an input device.
2. **Velocity-controlled joystick:** A return-to-zero joystick controls the velocity of an object (such as a cursor) that is continuously repositioned on the display. Zero displacement of the joystick corresponds to no motion (zero velocity). Positioning with a velocity-controlled joystick (a temporally and spatially constrained task) has been shown to be inferior to positioning with a mouse (a spatially constrained task) for desktop pointing scenarios (Card, English, & Burr, 1978). Silfverberg et al. (Silfverberg, MacKenzie, & Kauppinen, 2001) have done an in-depth study of isometric joysticks on handheld devices to control the cursor on a situated public display. Many of today's mobile phones are shipping with simple joysticks with a push button for menu navigation.

3. **Accelerometers:** Accelerometers are beginning to emerge in handheld devices. For example, Samsung's SCH-S310 mobile phone comes with an integrated 3-D accelerometer. Several researchers (Bartlett, 2000; Harrison et al., 1998; Hinckley and Horvitz, 2001) have proposed interactions that allow users to scroll (e.g., through an electronic photo album) by tilting the handheld device. The scrolling is typically activated through a clutch mechanism, such as squeezing the sides of the device (Harrison, Fishkin, Gujar, Mochon, & Want, 1998). The degree of tilting controls the speed of scrolling, making this a temporally constrained positioning task similar to the velocity-controlled joystick. Although these techniques were used to interact with an application directly on the device, they could clearly be extended to positioning tasks in ubiquitous computing environments.
4. **Camera tracking:** C-Blink (Miyaoku, Higashino, & Tonomura, 2004) rapidly changes the hue of a color phone screen to allow an external camera system to track the phone's absolute motion for cursor control on a large public display. The hue sequence encodes an ID to allow multiple users to interact simultaneously and control independent cursors.

   The Smart Laser Scanner uses a laser combined with a wide-angle photo detector (see Figure 1) to detect relative finger motion in 3-dimensional space (Cassinelli, Perrin, & Ishikawa, 2005). The laser beam is steered with a two-axis micro-mirror. The tracking principle is based on the backscatter of a laser beam. When the backscatter is disrupted, the motion is deduced from the angle of the backscatter, and the laser is repositioned for the next measurement. Like other tracking techniques, it is possible for the device to lose track if the finger moves too fast, but input can easily be resumed by repositioning the finger to the laser. The research prototype of the tracker is fast enough to track the motion of a bouncing ping-pong ball.
5. **Motion detection:** With the Sweep (Ballagas, Rohs, Sheridan, & Borchers, 2005) interaction technique, the phone is waved in the air to control relative cursor motion on a remote screen (see Figure 2). This is accomplished using motion detection, an image processing technique involving rapidly sampling successive images from the phone's camera and sequentially comparing them to determine relative motion in the (x, y, θ) dimensions. No visual tags are required. The screen on the phone can be ignored, and the camera does not even need to be pointed at the display. A

Figure 1. The Smart Laser Scanner: A 3-D input technique for mobile devices using laser tracking (Cassinelli et al., 2005). Reprinted with permission from the authors.



clutch mechanism, such as a button press, is used to activate the Sweep interaction. The clutch can be used to reposition the arm, similar to the way a mouse can be lifted to be repositioned without additional cursor motion.

6. **Location detection:** Location of the phone can also be used as input, where the user moves through physical space. Mogi (Licoppe & Inada, 2006), for instance, is a phone-based persistent item collection and trading game where the absolute geo-position of a subscriber correlates to the position in the game world. Mogi combines GPS (global positioning system) technology built into the phone with information from different mobile infrastructure towers from the network service provider to determine the player's position.

Continuous Direct Interactions

7. **Camera tracking:** Madhavapeddy, Scott, Sharp, and Upton (2004) present camera-based interactions involving tagging interactive GUI elements such as sliders and dials (see Figure 3). In manipulating the position and orientation of the phone camera, the user can position a graphical slider, or orient a graphical dial. Similarly, Direct Pointer (Jiang, Ofek, Moraveji, & Shi, 2006) uses a handheld camera to track the standard cursor on the display. An analogy can be drawn to the classic light pen with a tracking cross. As the light pen moves to a new position, the cross follows the motions of the pen. Tracking may be lost if the pen is moved too fast, but can be easily resumed by repositioning the pen back to the tracking cross. Madhavapeddy et al.'s interactions rely on the tagged GUI widget instead of a cross for tracking; in Direct Pointer, the mouse cursor is the modern equivalent of the tracking cross.

   In these tracking examples, the handheld device is responsible for tracking. An alternative is to use a tracker in the environment to track the output from a handheld device. For example, smart phones have been augmented with laser pointers, as in Patel and Abowd (2003), making them suitable for positioning tasks, described by Dan, Olsen, and Nielsen (2001), that use a camera in the environment to track the laser.

   The mobile phone can also be passively tracked using a camera in the environment, such as in VisionWand (Cao & Balakrishnan, 2003). The user holds a passive handheld device that is augmented with distinctive markings (such as colored balls) at each end. Using two fixed cameras to perform stereo tracking, a 3-D ray can be deduced from the orientation of the markings in the stereo view, assuming the distance of the markings on the device is known *a priori*. This allows using a projection of the ray as a pointing device for a fixed remote screen.

Figure 2. The Sweep technique uses camera input and optical flow image processing to control a cursor (Ballagas et al., 20*05). © 2006 IEEE. Adapted with permission.*

The result is an interaction that is very similar to pointing using a laser pointer, except the ray is not a visible beam of light. This technique has an advantage over the standard laser pointer in that it provides an extra dimension of information: the distance to the display. The disadvantage of this interaction is that it is vulnerable to occlusion (e.g., by the users' own body) bringing into question the robustness of tracking in practical scenarios, although different camera configurations (such as from overhead facing downward) may solve these issues for certain interaction scenarios.

Discrete Indirect Interactions

8. **Directional step keys:** The location of an object is controlled using up, down, left, and right step keys for 2-D applications, plus in and out for 3-D. In the Blinkenlights project (Chaos Computer Club, 2002), users played the arcade classic "Pong" using the side of a building as a large public display. Each window equaled one pixel on the 18x8 pixel display. Players connected to the display by making a standard voice call to a phone number. Pressing the number 5 on the phone keypad moved the paddle up, and the number 8 moved it down. The server controlling the "Pong" application would decode the tones generated from the key activity during the phone call and use them as application input. One of the notable things about this interaction is that it used the lowest common denominator of phone technologies. The communications channel was the standard voice channel, and the input was the numeric keypad, requiring no additional hardware or software besides what standard phones provide.

Discrete Direct Interactions

9. **Camera image:** Using the Point and Shoot (Ballagas et al., 2005) interaction technique, the user can specify an absolute position on a public display using a cross-hair drawn over a live camera image on the mobile phone. To make a selection, the user presses a button while aiming at the desired target.[1] The button press triggers a brief overlay of a grid of 2-D tags over the large display contents, as can be seen in the middle of Figure 4. The grid allows the phone to derive a perspective-independent coordinate system on the large display that is enabled by the special properties of the Visual Code tags (Rohs, 2005a). Only one visual tag is required to establish a coordinate system, but a grid is used to increase the probability of having one tag entirely in the camera view. The drawback of the current implementation is that the tag grid is disruptive in multiuser scenarios, but future implementations could, for example, display the tags in infrared so

Figure 3. Using the phone to manipulate tagged widgets such as buttons, dials, and sliders (Madhavapeddy et al., 2004). Reprinted with permission from the authors.

that they are visible to the camera but not to other users.

Point and Shoot is related to the classic light pen, where position is discretely determined by displaying a raster scan when the user clicks a button on the light pen. When the raster scan is temporally sensed by the pen, the position of the pen is known because of a tight coupling between the pen clock and display clock. In Point and Shoot, a visual tag grid replaces the functionality of the raster scan except its mechanics are spatial rather than temporal. The lack of temporal dependencies makes Point and Shoot robust to different display technologies and the loose coupling between camera and display.

The breadth of positioning techniques is relatively large, making it difficult to choose which technique is most appropriate for a particular application scenario. To help with this selection, it is important to examine different figures of merit for each device.

Evaluating Positioning Techniques

There have been only a handful of thorough evaluations of the different ubiquitous mobile input techniques (Ballagas et al., 2005; Myers, Bhatnagar, Nichols, Peck, Kong, Miller, & Long, 2002; Silfverberg et al., 2001; Wang, Zhai, & Canny, 2006), as the field is still relatively new. These studies are difficult to directly compare, since they each used different experimental parameters, and some evaluations were not done in the context of ubiquitous computing interaction scenarios. Therefore, rough estimates for a variety of ergonomic measures are used to create a high-level comparison table for the positioning task presented in Figure 6. These rough estimates are derived using our knowledge of the interaction techniques for mobile phones and the collective knowledge of their desktop computing counterparts. The ergonomic parameters are mostly borrowed from Foley et al.'s survey of interaction techniques.

The evaluation measures are grounded in psychological and physiological foundations. Card et al. (Card, Newell, & Moran, 1983) provide an integrated survey of the various fundamental theories in a way that makes them more accessible and easier to use during analysis. Central to this work is the human processor model, which brings knowledge of the perceptual, cognitive, and motor processes of a human together under a single model. Ideally, a user interface minimizes the work required for each of these basic psychological processes.

The comparison table also incorporates various ergonomic measures designed to capture the efficiency of users executing the subtask, the accuracy they can achieve, and the pleasure the user derives from the process. The individual measures used in our comparison table are as follows:

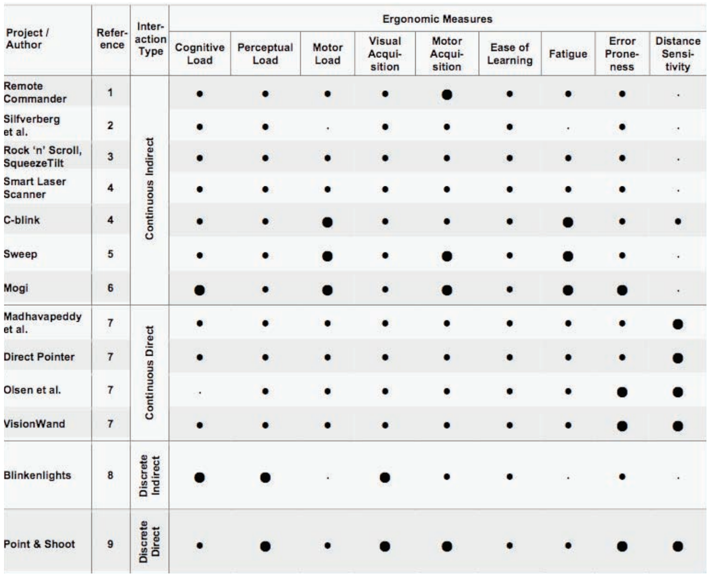
- **Perceptual load** refers to the difficulty for the user to recognize, with their own senses, the physical stimuli and feedback of the interaction. For example, in the Point and Shoot interaction, users need to shift their perceptual attention between a large display and the phone screen to isolate a target in the phone camera view, leading to a comparatively high perceptual load.
- **Cognitive load** refers to the difficulty for the users to organize and retrieve information related to the interaction technique.
- **Motor load** refers to the number of motor steps required to execute the action after the appropriate action has been determined in the cognitive process. For example, Mogi is classified as a high-motor-load technique because the user needs to physically move at the city scale to specify the necessary position.
- **Motor acquisition time** characterizes the amount of time for the processes involved in the interaction technique (i.e., reaching for an object, moving to a certain target area, rotating to a certain orientation, etc.)
- **Visual acquisition time** characterizes the amount of time it takes to perceive the physical stimuli of the interaction technique.

Figure 4. Point & Shoot interaction: (Left) The phone screen is used to aim at a puzzle piece on a large situated display. (Middle) Pressing the joystick indicates a selection and a Visual Code grid is briefly superimposed to compute the target coordinates in the captured photo. (Right) The grid has disappeared and the target puzzle piece is highlighted on the large display, indicating successful selection (Ballagas et al., 2005). © 2006 IEEE.

Figure 5. Summary of positioning techniques using a smart phone as an input device. © 1984, 2006 IEEE. Adapted by permission.

	Foley	Mobile Phone Interactions	
	In Environment	On Phone	In Environment
Direct Pick	**Search for Light Pen** (Raster Scan)	**Camera + On-Screen Cursor +**	**Visual Tags** [Point & Shoot]
		Laser Pointer +	**Camera** [Olsen]
	Light Pen Tracking	**Camera Tracking +**	**Visual Tags** [Madhavapeddy]
		Camera Tracking +	**On-Screen Cursor** [Direct Pointer]
		Vision Markers +	**Camera** [VisionWand]
		Camera + Touch-Screen	
Direct with Locator Device	**Touch Panel**		
Indirect with Locator Device	**Mouse**	**Camera** [Sweep]	
		Display +	**Stationary Camera** [C-blink]
	Joystick (Velocity)	**Joystick** (Velocity) [Silfverberg]	
	Tablet	**Trackpad** [Remote Commander]	
	Trackball		
	Cursor Control Keys with Auto-Repeat	**Cursor Control Keys with Auto-Repeat**	
	Joystick (absolute)	**Accelerometer** [Harrison + Rock 'n' Scroll]	
		Steerable Laser + Wide Angle Photodetector [Cassinelli]	
Indirect with Directional Commands/ Button Push	**Up-Down-Left-Right Arrow Keys**	**Directional Step Keys** [Blinkenlights]	
	(See Selection)	(See Selection)	
Location Detection		**GPS +**	**Cell Network Towers** [Mogi]
Numerical Value/ Numerical Coordinates/ Character String Name	(See Text Input)	(See Text Input)	

Figure 6. Rough estimates of ergonomic measures to compare mobile-phone-based positioning techniques (small circle = low, medium circle = medium, large circle = high). © 2006 IEEE. Adapted by permission.



- **Ease of learning:** Characterizes the level of skill that is required to use the device.
- *Fatigue* characterizes how tiring the interaction technique is to perform.
- **Error proneness:** Characterizes the susceptibility for errors of the input technique, the degree to which the interaction technique, by its design, allows/avoids errors, for example, if possible movement trajectories match the degrees of freedom of the required input then certain errors can be avoided.
- **Sensitivity to distance:** Users in ubiquitous computing scenarios typically have freedom of motion, making the amount of separation between the user and the target in the environment (such as a large display or other device) dynamic and unpredictable. Thus, the range of distances the interaction will support is an important design consideration. Interactions that are based on aiming, such as laser pointers, become more difficult to perform when further away, where targets are perspectively smaller. Other techniques, such as the Sweep technique, are not significantly affected by distance of interaction.

Orient

The ORIENT subtask involves specifying a heading or direction instead of a position. Like POSITION, ORIENTATION is also not limited to graphics subtasks as it can relate to physical orientation in the real world, such as a security camera, a spotlight, or a steerable projector. Some of Foley et al.'s original graphics interactions carry over directly to ubiquitous computing, including *indirect continuous orientation with velocity-controlled joystick* and *discrete orientation with angle type-in*. The remaining techniques observed in our survey include

Continuous Indirect Interactions

1. **Locator device:** The user can specify the angle of orientation by using a continuous quantifier or one axis of a positioning device. The Sweep technique supports detection of rotation around the Z-axis (perpendicular to the display), allowing interactions like rotating a puzzle piece in a jigsaw puzzle

application, where the phone is used like a ratchet to adjust orientation. The image processing used by Sweep also detects rotation around the X and Y-axis. However, for better performance as a positioning device, rotation around the Y-axis is mapped to translation along the X-axis, and rotation around the X-axis is mapped to translation along the Y-axis.

2. **Camera tracking:** VisionWand (Cao & Balakrishnan, 2003) uses a set of cameras in the environment to track the absolute orientation of a marked handheld device. The technique requires that at least two markers are visible in at least two camera viewpoints to determine the orientation in 3-dimensional space.

Continuous Direct Interactions

3. **Camera tracking:** Madhavapeddy's tagged GUI dials (Madhavapeddy et al., 2004) can be oriented using the phone camera to track rotation movement. Similar to the Sweep technique, the phone is used like a rachet to adjust orientation.
4. **Compass:** Electronic compasses, such as the Honeywell HMC1052 magnetometer, can be used to detect the physical orientation of the phone with a +/-3° error, enabling a continuous and direct ORIENT task. This or similar sensors could be easily incorporated into future mobile phone applications.

Discrete Direct Interactions

5. **Camera image:** The Point & Shoot technique supports discrete orientation along the Z-axis. As the user aims at a target, they rotate the phone to specify the desired Z-orientation using the aiming cross-hair as an axis of rotation.

Select

In many interaction scenarios, the user must choose from a set of alternatives, such as a menu of icons. The SELECTION subtask addresses this style of interaction. The SELECTION subtask is commonly accomplished by arranging the items spatially in a graphical user interface, allowing the user to complete the selection using a cursor controlled through the positioning subtask. Instead of icons, the set of alternatives might be a list of commands. However, selection is not limited to graphical interactions, as a user may select a physical object

Figure 7. Summary of ORIENT techniques using a smart phone as an input device. © 1984, 2006 IEEE. Adapted by permission.

	Foley	Mobile Phone Interactions	
	In Environment	On Phone	In Environment
Direct Pick		**Camera** [Point & Shoot]	
		Camera Tag Tracking +	**Visual Tags** [Madhavapeddy]
		Vision Markers +	**Camera** [VisionWand]
Indirect with Cursor Match/ Locator Device		**Camera** [Sweep]	
	Joystick (Velocity)	**Joystick** (Velocity)	
	Joystick (absolute)		
Numerical Value/ Numerical Coordinates/ with Character String Name	(See Text Input)	(See Text Input)	

to operate upon, such as selecting a lamp to adjust its setting. Many selection techniques carry over directly from Foley et al.'s earlier analysis, such as *character string name type-in* common for command prompts, or *button push–soft keys,* where buttons are located on the edge of the display area with their labels displayed on screen. The remaining selection techniques are as follows:

Continuous Indirect Interactions

1. **Gesture recognition:** The user makes a sequence of movements with a continuous positioning device such as the joystick, camera, trackpad, or accelerometers. For example, Patel, Pierce, & Abowd (2004) used gesture recognition of accelerometer data from the handheld device to authenticate users that wanted to access data on their mobile phone through an untrusted public terminal. Using this technology, users could securely bring up data on the public terminal from their phone without removing it from their purse.

Continuous Direct Interactions

2. **Tagged objects:** RFIG Lamps (Raskar, Beardsley, van Baar, Wang, Dietz, Lee, Leigh, & Willwacher, 2004) allows a handheld projector to be used to select objects with photosensitive RFID tags in the physical world. The handheld projector emits a graycode pattern that allows the tags to determine their relative position in the projected view. Waving the handheld projector around, you can navigate a cursor in the center of the projected view to select individual physical objects.

Discrete Indirect Interactions

3. **Voice recognition:** The user speaks the name of the selected command, and a speech recognizer determines which command was spoken. The Personal Universal Controller (Nichols & Myers, 2006) supports automatic generation of speech interfaces (as well as graphical interfaces) to issue commands to objects in the real world. VisionWand (Cao & Balakrishnan, 2003) also demonstrates a rich gesture vocabulary using stereovision to track a passive wand. For example, a tapping gesture is used to allow selection of the current cursor position specified by the orientation of the wand. It should be noted that information from any continuous positioning technique can be used for gesture recognition, as long as there is a mechanism to specify when a gesture begins and ends.

Discrete Direct Interactions

4. **Tagged objects:** Tagged objects can be used to present information on a wireless mobile computer equipped with an electronic tag reader, as demonstrated by the early E-tag project (Want, Fishkin, Gujar, & Harrison, 1999). For example, selecting a book by scanning its embedded RFID tag would activate a virtual representation of the object on the screen, such as a Web reference to the book allowing it to be purchased. Similar interactions have also been proposed for visual tags in the environment (Rohs, 2005a) and tagged GUI elements (Madhavapeddy et al., 2004; Rohs, 2005b), where a camera is used to acquire an image to decode the selected tag. Patel and Abowd (2003) present a physical world selection method for mobile phones in which a modulated laser pointer signal triggers a photosensitive tag placed in the environment, allowing users to bring up a menu to control the object on their handheld device.
5. **Laser pointer:** Myers et al. (2002) proposed a multilayer selection technique, called "semantic snarfing," that combines multiple devices in consecutive actions. First, a laser pointer integrated with a handheld computer is used to make a coarse-grained selection of a screen region on a display in the environment. A camera, also in the environment, detects

Figure 8. Summary of SELECTION *techniques using a smart phone as an input device (Continued on next page). © 1984, 2006 IEEE. Adapted by permission.*

	Foley	Mobile Phone Interaction	
	In Environment	On Phone	In Environment
Direct Pick		**Camera + On-Screen Cursor** [Point & Shoot]	
		Laser Pointer + [Olsen, Semantic Snarfing, Patel]	**Light Sensor** (e.g. camera)
	Light Pen Tracking	**Camera +** [Madhavapeddy] **Camera Tracking +** [Direct Pointer] **Camera +**	**Visual Tags** **On-Screen Cursor** **Pen Input**
		Vision Markers + [VisionWand]	**Camera + Tapping** **Gesture Recognition**
		Handheld Projector + **RFID Reader** [RFIG]	**Light Sensitive** **RFID Tags**
		RFID Reader + [Want]	**RFID Tags**
	Touch Panel		
Indirect with Cursor Match/ Locator Device	**Mouse**	**Camera** [Sweep]	
		Display + [C-blink]	**Stationary Camera**
	Joystick (Velocity)	**Joystick** (Velocity) [Silfverberg]	
	Tablet	**Trackpad** [Remote Commander]	
	Trackball		
	Cursor Control Keys	**Cursor Control Keys**	
	Joystick (absolute)	**Accelerometer** [Harrison, Rock 'n' Scroll]	
		Steerable Laser + **Wide Angle Photodetector +** **Button Push** [Cassinelli]	

Figure 9. Summary of SELECTION *techniques using a smart phone as an input device (Continued from previous page). © 1984, 2006 IEEE. Adapted by permission.*

	Foley	Mobile Phone Interaction	
	In Environment	On Phone	In Environment
Indirect with Directional Commands/ Button Push/ Time Scan	**Programmed Function Keyboard**	**Programmed Function Keyboard**	
	Soft Keys	**Soft Keys**	
	Alphanumeric Keyboard		
Gesture Recognition/ Sketch Recognition		**Camera + Button Push Clutch** [TinyMotion]	
		Display +	**Camera**
		Accelerometer [Patel]	
	Light Pen	**Pen Input**	
		Steerable Laser + Wide Angle Photodetector + Button Push for Clutch [Cassinelli]	
		Laser Pointer + [Olsen, Semantic	**Light Sensor** (e.g. camera) Snarfing, Patel]
		Vision Markers + [Vision	**Camera** Wand]
	Tablet + Stylus		
Location Detection		**GPS +**	**Cell Networks Towers** [Mogi]
Voice Input		**Microphone +** [PUC]	**Voice Recognizer**
Numerical Value/ Numerical Coordinates/ with Character String Name	(See Text Input)	(See Text Input)	

laser activity on the display. The system then transmits the details of the selected screen region to the handheld device, which composes a GUI on the handheld screen to make the fine-grained selection with a stylus.

Path

The PATH subtask involves specifying a series of positions and orientations over time. The PATH subtask has different requirements than POSITION and ORIENT because the movement is governed by the speed-accuracy tradeoff (Schmidt, Hawkins, Frank, & Quinn, 1979). Despite this, PATH adheres to the same taxonomy as the corresponding POSITION and ORIENT techniques, because a PATH task can be specified using the more primitive subtasks.

Quantify

The QUANTIFY task involves specifying a value or number within a range of numbers. This technique is used to specify numeric parameters such as time or speaker volume. In ubiquitous applications, QUANTIFY tasks using phone input were typically accomplished through the GUI using 1-D POSITION or ORIENT subtasks.

Text

TEXT ENTRY for mobile phones is a well-studied area (MacKenzie & Soukoreff, 2002) as it is central to text-based mobile communications like SMS (short messaging service) and personal information management functionality. Text entry also has many applications for ubiquitous applications, for example, the Digital Graffiti (Carter, Churchill, Denoue, Helfman, & Nelson, 2004) project seeks to annotate public content on large public displays. This section is not intended to be a comprehensive survey of mobile text entry techniques, but we have selected a few examples to illustrate the design space. All of the techniques listed were originally designed for text input directly on the mobile phone, but could clearly be used for text entry for a ubiquitous computing application.

Keyboard

Although some mobile phones and handheld devices feature a full QWERTY keyboard (albeit much smaller than their desktop counterparts), miniaturization trends make this type of keyboard impractical in a majority of mobile phone form factors. The most well known text entry techniques for mobile phones use a standard numeric keypad. For text entry from a 26-character alphabet using this keyboard, a mapping with more than one character per button is required. Following the classification by Wigdor and Balakrishnan (2004), there are two fundamental types of disambiguation: *consecutive*, where the user first selects a letter grouping and then an individual letter, or *concurrent*, where the user simultaneously selects the letter grouping and the individual letter.

Consecutive approaches are the most common today. One approach to disambiguate text entry is MultiTap, which requires users to make multiple presses to select a single letter from the characters associated with a certain key. Another solution is to use a two-key disambiguation, where the first key selects the letter group, and the second key specifies the letter in the group. Dictionary-based techniques, such as T9[2], deduce the word being typed, based on the different possibilities for combining the groups of characters. When multiple words match the key sequence, the user selects the intended word from a list (typically ordered by probability or frequency of use).

Concurrent approaches, however, demonstrate a lot of promise. For example, TiltText (Wigdor & Balakrishnan, 2003) combines the standard 12-key keypad with an accelerometer. To disambiguate which character is intended when a key is pressed, TiltText uses the tilt orientation of the handset. A keypress with the phone tilted to the left enters the first character on the key, forward tilt enters the second character, right tilt enters the third character, tilting towards the user enters the fourth character (if one exists for the key), and no tilt enters the numeric character.

ChordTap (Wigdor & Balakrishnan, 2004) combines the standard numeric keyboard with additional "chording" buttons on the back of the phone. A user selects an individual letter by selecting the key group on the numeric keyboard while pressing the appropriate "chord" key on the back of the phone.

If miniaturization trends continue, TiltType (Partridge, Chatterjee, Sazawal, Borriello, & Want, 2002) represents an interesting point in the design space that combines chord button presses to specify a letter grouping and tilting to allow the user to specify a particular character within that grouping. Using only four buttons and a two-axis accelerometer, the technique supports an alphabet of 55 characters in a watch-sized form factor. Expert users can memorize the character positions, allowing the letter grouping and individual character within the grouping to be specified concurrently.

Speech Recognition

Text entry by speech recognition is not yet technically viable on mobile platforms, but we list it here for completeness. Technology is making rapid advances in the realm of speech processing. For example, system on a chip designs for speech processing (Ravindran, Smith, Graham, Duangudom, Anderson, & Hasler, 2005) have the potential to bring speech input to interactive text entry on mobile phones. Karpov et al. (Karpov, Kiss, Leppanen, Olsen, Oria, Sivadas, & Tian, 2006) have developed a short message (SMS) dictation system for Symbian phones with a vocabulary of 23,000 words. The language model is adapted to words typically used in SMS messages.

Speech recognition could also be achieved in a compound architecture where the speech is recognized through an external computer (i.e., connected through a voice call) and sent back to the mobile phone.

Stroked Character Recognition

Pen-based techniques, such as Graffiti, are very common in the PDA form factor, and are also available on a small portion of the handsets on today's market. However, any of the continuous positioning tasks discussed earlier are capable of generating stroke information necessary for stroked-character recognition. For example, TinyMotion (Wang et al., 2006) demonstrates both English and Chinese stroked character recognition using camera-based motion estimation (similar to the Sweep technique).

Menu Selection

On-screen keyboards are common for touch sensitive displays, where the letters of the alphabet are displayed as a menu of buttons, commonly in a spatial layout similar to the QWERTY keyboard. If the screen size of the mobile phone is not large enough to depict a keyboard layout, items in the environment could be used to display the menu, where users select the characters using the SELECTION subtask previously discussed.

Figure 10. Summary of TEXT ENTRY techniques using a smart phone as an input device. © 1984 IEEE. Adapted by permission.

	Foley	Mobile Phone Interactions	
	In Environment	On Phone	In Environment
Keyboard	**Alphanumeric**	**Alphanumeric** [Multitap, T9]	
		Alphanumeric + Accelerometer [TiltText]	
	Chord	**Alphanumeric + Chord** [ChordTap]	
		Chord + Accelerometer [TiltType]	
Stroked Character Recognition	**Tablet with Stylus**	(See Continuous Positioning)	
Voice/Speech Recognition	**Voice Recognizer**	**Speech Recognizer**	
		Microphone +	**Speech Recognizer**
Direct Pick from Menu with Locator Device	**Light Pen**	(See Selection)	
	Touch Panel	(See Selection)	
Indirect Pick from Menu with Locator Device	(See Positioning)	(See Positioning)	

SPATIAL LAYOUT OF THE DESIGN SPACE

Our interaction taxonomy is summarized in Foley-style graphs in Figures 5, 7, 8, 9, and 10. Card et al. (1991) point out that this format is somewhat ad hoc and lacks a notion of completeness. Card then builds on the work of Buxton (1983) to create a systematic spatial layout of the design space of input devices that captures the physical properties of manual devices very well. However, it does not capture many aspects that are relevant to ubicomp interactions such as modality or feedback (Ballagas et al., 2003).

Using Foley et al.'s taxonomy, we propose a five-part spatial layout, shown in Figure 11, for mobile phone interaction tasks discussed in our survey including supported subtasks (Position, Orient, and selection), dimensionality, relative vs. absolute movement, interaction style (direct vs. indirect), and feedback from the environment (continuous vs. discrete). Feedback and interaction style have been previously defined in the introduction to Foley et al.'s taxonomy. We describe the remaining dimensions in more detail in the remainder of this section.

Supported Subtasks

When choosing the most appropriate input device for a particular interaction scenario, the subtasks an interaction supports are the primary consideration. By including the subtask directly in the design space, it becomes more useful as a design tool.

Dimensionality

Dimensionality refers to the number of dimensions the interaction supports. Dimensionality can indicate spatial dimensions (X,Y,Z) or rotational dimensions (rX,rY,rZ). This distinction is visible in our design space by observing the subtask of the dimension. Following Card et al. (1991), if a particular interaction uses a combination of dimensions across different points in the design space, the relationship is indicated using a merge composition operator (a solid line). In contrast to Card's notation, our merge composition operators are connecting subtasks, not spatial sensor dimensions.

Relative vs. Absolute

Relative input is specified with respect to interaction history: the input technique provides information about the amount of change from the previous state. Relative input can be specified regardless of the current physical properties, such as position and orientation. For example, standard desktop mouse input is specified through motion across the desktop regardless of the physical position of the mouse on the desktop.

Absolute input is specified with respect to current physical properties, and can be specified independently of any interaction history. For example, stylus input can be used to provide absolute positional information on a screen space.

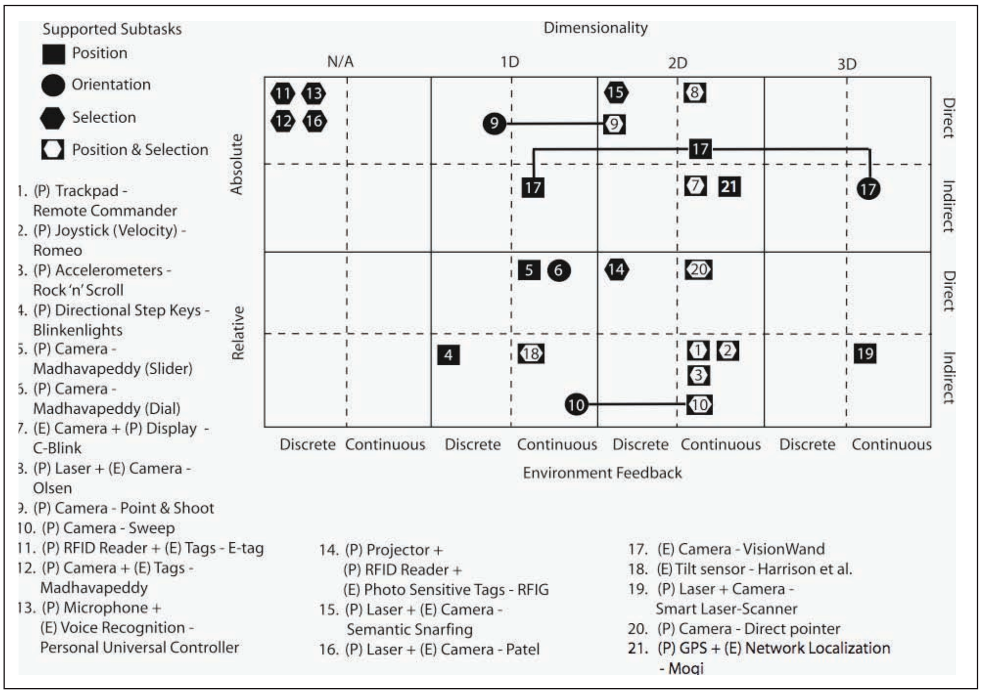
Other Relevant Attributes of Interaction Devices

It should be noted that this set of dimensions is not comprehensive, and other dimensions, such as resolution, direction (input vs. output), and modality, may provide further insights into the design space. However, the design space depicted in Figure 11 does provide an interesting overview of the interaction techniques covered in this chapter. Using this graphical layout, we are able to pinpoint gaps in the breadth of the interaction techniques surveyed, and can anticipate opportunities for future work. For example, our space shows no interaction that supports 3-dimensional relative direct orientation. An alternative layout might include direction and modality, which would demonstrate the sparse usage of auditory and haptic feedback in these techniques.

Designing for Serendipity

One key design consideration is the ease and speed of setting up a data connection between the phone and the environment or the device it is controlling. In some of the interactions surveyed,

Figure 11. Classification of different mobile phone interactions that have been implemented in the projects surveyed. Inspection of the diagram reveals opportunities for future work, for instance, developing interaction techniques that support 3-D relative direct orientation. In the listing of techniques, (P) indicates capabilities of the phone, and (E) indicates capabilities of the environment. © 2006 IEEE. Adapted by permission.



the data connection is inherent in the physical properties of the device. The VisionWand, for example, is a completely passive system, so it requires no additional action on the user's part to start the interaction.

The C-Blink interaction is classified as highly serendipitous, as the users merely launch an application on their mobile phone to interact with a display; no network connection or handshaking is required. The RFIG Lamps project also falls into this category because RFID tags are so simple in terms of communications protocol that no connection need to be established before data can be transferred.

For projects that use short-range wireless communications models, such as Bluetooth, visual or RFID tags can be used to encode the connection information for the environment, creating a very low threshold of use.

Social Acceptance

Smart phones today are social devices. While smart phone ubiquity seems inevitable, social acceptance will influence the success of these new interactions. Remind yourself, for example, of the first time you came across a person using a wireless headset to communicate via their mobile phone. For many people, this communication technique is still awkward and strange, particularly in public places. Smart phone interaction will require users to perform particular actions and behaviors that might feel unintuitive and awkward to them. Furthermore, they will perform these actions in the presence of passive or active others, both familiars and strangers. On one hand, outside observers might find these interactions disturbing or embarrassing, but on the other hand, these kinds of interaction have the potential to raise your social status, similar to the way phones themselves are status symbols for part of our society.

DESIGN SPACES IN THE DESIGN PROCESS

Design spaces are a particularly useful design tool as a part of a human-centered iterative design process (Nielsen, 1993). One of the pitfalls of iterative human-centered design is that if you pick a poor starting point, you may reach a peak in the usability of a particular design without reaching your desired usability goals. In this case, it may be necessary to throw the design away and start over. False starts are relatively painless early in the design process, but can be extremely expensive if determined late in the design process. In order to minimize the risk of false starts, a parallel design strategy (Nielsen & Faber, 1996) can be used, where multiple designs can be explored independently early in the design process. As the designs mature, the best design becomes clear, or the strengths of the top designs can be merged to a unified design. Using the design space, designers can more easily reason about alternative input techniques in a parallel design process.

As a concrete example, REXplorer (Ballagas, Walz, Kratz, Fuhr, Yu, Tann, Borchers, & Hovestadt, 2007) is a pervasive spell-casting game that allows tourists to explore the history of the medieval buildings in Regensburg, Germany. The game premise is that historical spirits are trapped inside of medieval buildings. Players need to interact with the spirits to learn their stories and perform quests on their behalf to earn points in the game. The game design called for spell-casting as the primary interaction metaphor; in order to awaken a spirit, one of four spells must be cast.

Choosing one spell out of four can be characterized as a SELECT subtask. The design space was used to identify a set of design alternatives that we initially considered:

1. Four dedicated spell buttons
2. Selecting one of four spells from on-screen menu
3. Recognition of spell gestures. We noted that gestures are actually specified using the path subtask. Then we came up with gesture input alternatives including:
    a. Pen trace across a touch screen
    b. Path using camera-based motion detection to allow the phone to be used like a magic wand.

After preliminary analysis with our target group (students aged 15-25), we decided to go with the

Figure 12. REXplorer uses the Sweep technique to allow players to cast spells using the PATH subtask (Ballagas et al., 2007). Reprinted with permission from the authors.

camera-based motion detection solution (see Figure 12). Waving the phone through the air is not the most efficient technique, but is the most similar to the spell-casting metaphor. Also, this physical style of gesture was more likely to create an engaging experience (Hummels, 2000).

Later in the design process, after a working-gesture recognition system was created, we did a full playability test. Most of the test players found the gestures to be an important element of gameplay. They found it heightened the sense of magic and mysteriousness. However, we also discovered during the playability tests that a few of our players (especially our older participants) found the gestures awkward. As a compromise, we created a unified design, where an alternative gesture selection mechanism through an on-screen menu can be used anytime an invalid gesture is performed, effectively allowing people to avoid gestures altogether if desired. This final design encouraged the use of gestures for spell selection to promote engagement, but allowed an alternative selection mechanism to those who preferred to avoid gestures.

CONCLUSION

Our structured tour illustrates the state of the art in using smart phones to interact with and control our environments. The taxonomy organizes the range of techniques into families that help make functional relations between the mobile phone techniques and their desktop counterparts. The design space addresses the lack of a sense of completeness in the taxonomy, and structures the range of interactions in a way that helps visually identify gaps and predict future interaction techniques. The design space can be used as a part of a human-centered iterative design process to help generate parallel or alternative designs. These methods of thought are intended to inspire new applications that use the mobile phone for interaction with the environment, as well as inform the design of future smart phone interaction techniques.

REFERENCES

Ballagas, R., Ringel, M., Stone, M., & Borchers, J. (2003). iStuff: A physical user interface toolkit for ubiquitous computing environments. In *CHI '03 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 537–544). New York, NY: ACM Press.

Ballagas, R., Rohs, M., Sheridan, J. G., & Borchers, J. (2005). Sweep and point & shoot: Phonecam-based interactions for large public displays. In *CHI '05 Extended abstracts of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1200–1203). New York, NY: ACM Press.

Ballagas, R., Rohs, M., Sheridan, J., & Borchers, J. (2006). The smart phone: A ubiquitous input device. *IEEE Pervasive Computing, 5*(1), 70–77.

Ballagas, R., Walz, S. P., Kratz, S., Fuhr, C., Yu, E., Tann, M., Borchers, J., & Hovestadt, L. (2007). REXplorer: A mobile, pervasive spell-casting game for tourists. In *To appear in CHI '07 Extended Abstracts on Human Factors in Computing Systems.* San Jose, CA: ACM Press.

Bartlett, J. F. (2000). Rock 'n' scroll is here to stay. *IEEE Comput. Graph. Appl., 20*(3), 40–45.

Buxton, W. (1983). Lexical and pragmatic considerations of input structures. *SIGGRAPH Comput. Graph., 17*(1), 31–37.

Cao, X. & Balakrishnan, R. (2003). VisionWand: Interaction techniques for large displays using a passive wand tracked in 3-D. In *Proceedings of the 16th annual ACM Symposium on User Interface Software and Technology* (pp. 173–182). ACM Press.

Card, S. K., English, W. K., & Burr, B. J. (1978). Evaluation of mouse, rate controlled isometric joystick, step keys and text keys for text selection on a CRT. *Ergonomics, 21*, 601–613.

Card, S. K., Mackinlay, J. D., & Robertson, G. G. (1991). A morphological analysis of the design space of input devices. *ACM Transactions Information Systems, 9*(2), 99–122.

Card, S. K., Newell, A., & Moran, T. P. (1983). *The psychology of human-computer interaction*. Lawrence Erlbaum Associates, Inc.

Carter, S., Churchill, E., Denoue, L., Helfman, J., & Nelson, L. (2004). Digital graffiti: Public annotation of multimedia content. In *CHI '04: Extended abstracts of the SIGCHI Conference on Human Factors and Computing Systems* (pp. 1207–1210). ACM Press.

Cassinelli, A., Perrin, S., & Ishikawa, M. (2005). Smart laser-scanner for 3-D human-machine interface. *CHI '05: Extended abstracts of the SIGCHI Conference on Human Factors and Computing Systems* (pp. 1138–1139).

Chaos Computer Club. (2002). *Blinkenlights*. Retrieved from http://www.blinkenlights.de

Dan, R. Olsen, J., & Nielsen, T. (2001). Laser pointer interaction. In *CHI '01 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 17–22). New York, NY: ACM Press.

Foley, J. D., Wallace, V. L., & Chan, P. (1984). The human factors of computer graphics interaction techniques. *IEEE Computer Graphics and Applications*, *4*(11), 13–48.

Harrison, B. L., Fishkin, K. P., Gujar, A., Mochon, C., & Want, R. (1998). Squeeze me, hold me, tilt me! An exploration of manipulative user interfaces. In *CHI '98 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 17–24). New York, NY: ACM Press/Addison-Wesley Publishing Co.

Hinckley, K., & Horvitz, E. (2001). Toward more sensitive mobile phones. In *UIST '01 Proceedings of the 14th annual ACM Symposium on User Interface Software and Technology* (pp. 191–192). New York, NY: ACM Press.

Hummels, C. C. M. (2000). *An exploratory expedition to create engaging experiences through gestural jam sessions*. Ph.D. thesis, Delft University of Technology.

Jiang, H., Ofek, E., Moraveji, N., & Shi, Y. (2006). Direct pointer: direct manipulation for large-display interaction using handheld cameras. In *CHI '06 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1107–1110). New York, NY: ACM Press.

Karpov, E., Kiss, I., Leppanen, J., Olsen, J., Oria, D., Sivadas, S., & Tian, J. (2006). Short message dictation on Symbian Series 60 mobile phones. In *ICMI '06 Proceedings of the 8th international conference on Multimodal interfaces* (pp. 126–127). New York, NY: ACM Press.

Licoppe, C., & Inada, Y. (2006). Emergent uses of a multiplayer location-aware mobile game: The interactional consequences of mediated encounters. *Mobilities*, *1*(1), 39–61.

MacKenzie, I., & Soukoreff, R. (2002). Text entry for mobile computing: Models and methods, theory and practice. *Human-Computer Interaction*, *17*(2), 147–198.

Madhavapeddy, A., Scott, D., Sharp, R., & Upton, E. (2004). Using camera-phones to enhance human-computer interaction. In *Sixth International Conference on Ubiquitous Computing (Adjunct Proceedings: Demos)*. Springer-Verlag.

Miyaoku, K., Higashino, S., & Tonomura, Y. (2004). C-blink: A hue-difference-based light signal marker for large screen interaction via any mobile terminal. In *UIST '04 Proceedings of the 17th annual ACM Symposium on User Interface Software and Technology* (pp. 147–156). ACM Press.

Myers, B. A., Bhatnagar, R., Nichols, J., Peck, C. H., Kong, D., Miller, R., & Long, A. C. (2002). Interacting at a distance: measuring the performance of laser pointers and other devices. In *CHI '02 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 33–40). New York, NY: ACM Press.

Myers, B. A., Stiel, H., & Gargiulo, R. (1998). Collaboration using multiple PDAs connected to a PC. In *Proceedings of the 1998 ACM Conference on Computer Supported Cooperative Work* (pp. 285–294). ACM Press.

Nichols, J., & Myers, B. A. (2006). Controlling home and office appliances with smart phones. *IEEE Pervasive Computing*, *5*(3), 60–67.

Nielsen, J. (1993). Iterative user-interface design. *Computer*, *26*(11), 32–41.

Nielsen, J., & Faber, J. (1996). Improving system usability through parallel design. *Computer*, *29*(2), 29–35.

O'Hara, K., Perry, M., & Churchill, E. (2004). *Public and situated displays: Social and interactional aspects of shared display technologies (Cooperative Work, 2)*. Norwell, MA: Kluwer Academic Publishers.

Partridge, K., Chatterjee, S., Sazawal, V., Borriello, G., & Want, R. (2002). TiltType: Accelerometer-supported text entry for very small devices. In *UIST '02: Proceedings of the 15th annual ACM Symposium on User Interface Software and Technology* (pp. 201–204). New York, NY: ACM Press.

Patel, S. N., & Abowd, G. D. (2003). A 2-way laser-assisted selection scheme for handhelds in a physical environment. In *Ubicomp '03 Proceedings of the 5th International Conference on Ubiquitous Computing* (pp. 200–207). LNCS 2864, Springer.

Patel, S. N., Pierce, J. S., & Abowd, G. D. (2004). A gesture-based authentication scheme for untrusted public terminals. In *Proceedings of the 17th annual ACM Symposium on User Interface Software and Technology* (pp. 157–160). ACM Press.

Raskar, R., Beardsley, P., van Baar, J., Wang, Y., Dietz, P., Lee, J., Leigh, D., & Willwacher, T. (2004). RFIG lamps: interacting with a self-describing world via photosensing wireless tags and projectors. *ACM Transactions on Graphics*, *23*(3), 406–415.

Ravindran, S., Smith, P., Graham, D., Duangudom, V., Anderson, D., & Hasler, P. (2005). Towards low-power on-chip auditory processing. *EURASIP Journal on Applied Signal Processing*, *7*, 1082–1092.

Rohs, M. (2005a). Real-world interaction with camera phones. In H. Murakami, H. Nakashima, H. Tokuda, & M. Yasumura (Eds.), *Second International Symposium on Ubiquitous Computing Systems (UCS 2004), Revised Selected Papers* (pp. 74–89), Tokyo, Japan. LNCS 3598, Springer.

Rohs, M. (2005b). Visual code widgets for marker-based interaction. In *IWSAWC'05 Proceedings of the 25th IEEE International Conference on Distributed Computing Systems – Workshops (ICDCS 2005 Workshops* (pp. 506–513). Columbus, OH.

Schmidt, R. A., Zelaznik, H. N., Hawkins, B., Frank, J. S., & Quinn, J. T. (1979). Motor-output variability: A theory for the accuracy of rapid motor acts. *Psychological Review*, *86*, 415–451.

Silfverberg, M., MacKenzie, I. S., & Kauppinen, T. (2001). An isometric joystick as a pointing device for handheld information terminals. In B. Watson & J. W. Buchanan (Eds.), *Proceedings of Graphics Interface 2001* (pp. 119–126). Canadian Information Processing Society.

Wang, J., Zhai, S., & Canny, J. (2006). Camera phone based motion sensing: Interaction techniques, applications and performance study. In *UIST '06 Proceedings of the 19th annual ACM Symposium on User Interface Software and Technology* (pp. 101–110). New York, NY: ACM Press.

Want, R., Fishkin, K. P., Gujar, A., & Harrison, B. L. (1999). Bridging physical and virtual worlds with electronic tags. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 370–377). ACM Press.

Weiser, M. (1991). The computer for the 21st century. *Scientific American*, *265*, 94–104.

Wigdor, D., & Balakrishnan, R. (2003). TiltText: Using tilt for text input to mobile phones. In *UIST '03 Proceedings of the 16th annual ACM Symposium on User Interface Software and Technology* (pp. 81–90). New York, NY: ACM Press.

Wigdor, D., & Balakrishnan, R. (2004). A comparison of consecutive and concurrent input text entry techniques for mobile phones. In *CHI '04 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 81–88). New York, NY: ACM Press.

Zwicky, F. (1967). The morphological approach to discovery, invention, research and construction. In *New Methods of Thought and Procedure* (pp. 273–297). Berlin: Springer.

KEY TERMS

Continuous Interaction: Interactions with a closed-loop feedback, where the user continuously gets informed of the interaction progress as the task is being performed.

Design Space: Design spaces provide a formal or semiformal way of describing and classifying entities along different dimensions, each listing relevant categories or criteria.

Direct Interaction: Input actions are physically coupled with the user-perceivable entity being manipulated (such as an image on a display). To the user, this appears as if there is no mediation, translation, or adaptation between input and output. Physical coupling can be achieved when the feedback spatially coincides with the input action, or at a distance if the user is manipulating a 3-D ray (such as with a laser pointer) that intersects directly with the entity being manipulated.

Discrete Interaction: Interactions with an open-loop feedback, where the user is only informed of the interaction progress after the task is complete.

Indirect Interaction: User activity and feedback occur in disjoint spaces, where scaling and abstraction between input actions and feedback are often necessary.

Input Technique: A specific way of providing data input to a computer through a combination of input devices and software for visual, auditory, or haptic feedback.

ENDNOTES

1. An alternative implementation of the Point & Shoot technique could use pen input instead of the cross-hair image so that the user repositions the cursor by selecting the desired position directly on the live camera image displayed on the phone screen.
2. www.tegic.com

Chapter XXV
Text Entry

Mark David Dunlop
University of Strathclyde, UK

Michelle Montgomery Masters
University of Strathclyde, UK

ABSTRACT

Text entry on mobile devices (e.g., phones and PDAs) has been a research challenge since devices shrank below laptop size: mobile devices are simply too small to have a traditional full-size keyboard. There has been a profusion of research into text-entry techniques for smaller keyboards and stylus input: some of which have become mainstream, while others have not lived up to early expectations. This chapter will review the range of input techniques, together with evaluations, that have taken place to assess their validity: from theoretical modelling through to formal usability experiments. Finally, the chapter will discuss criteria for acceptance of new techniques, and how market perceptions can overrule laboratory successes.

INTRODUCTION

Although phones have traditionally been used for voice calls, with no need for text entry, many services such as text messaging, instant messaging, e-mail and diary operations require users to be able to enter text on phones; text messaging has even overtaken voice calling as the dominant use of mobile phones for many users. Phones and palmtop computers (or electronic organisers/personal digital assistants, PDAs) are too small for a standard desktop or laptop keyboard, thus requiring miniaturisation of the input methods. Furthermore, handheld screen technologies are making it increasingly convenient to read complex messages or documents on handhelds, and cellular data network speeds are now often in excess of traditional wired modems and considerably higher in wi-fi hotspots. These technological developments are leading to increased pressure from users to be able to author complex messages and small documents on their handhelds. Researchers in academia and industry have been working since the emergence of handheld technologies for new text-entry methods that are small and fast but easy-to-use, particularly for novice users. This chapter will look at different approaches to keyboards, different approaches to stylus-based entry, and how these approaches have been evaluated to establish which techniques are actually faster or less error prone. The focus of the chapter is both to give a perspective on the breadth of research in text entry, and also to look at how researchers have evaluated their work. Finally, we will look at perceived future directions, attempting to learn from the successes and failures of text-entry research.

KEYBOARDS

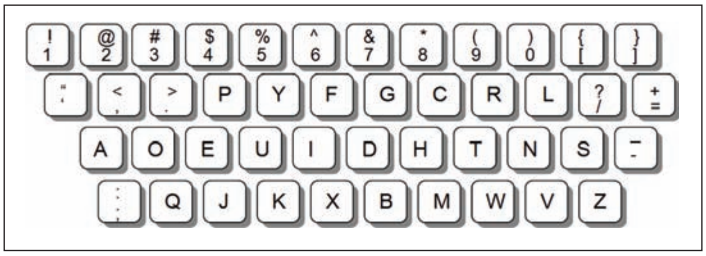
The simplest and most common form of text entry on small devices, as with large devices, is a keyboard. Several small keyboard layouts have been researched that try to balance small size of overall device against usability. These approaches can be categorized as unambiguous, where one key-press unambiguously relates to one character, or ambiguous, where each key is related to many letters (e.g., the standard 12-key phone pad layout where, say, *2* is mapped to *ABC*). Ambiguous keyboards rely on a disambiguation method that can be manually driven by the user or semiautomatic with software support and user correction. This section looks first at unambiguous mobile keyboard designs, then at ambiguous designs and, finally, discusses approaches to disambiguation for ambiguous keyboards.

Unambiguous Keyboards

Small physical keyboards have been used in mobile devices from their very early days on devices such as the Psion Organiser in 1984 and the Sharp Wizard in 1989, and have seen a recent resurgence in devices targeting e-mail users, such as most of RIM's Blackberry range. While early devices tended to have an alphabetic layout, the standard desktop layout was soon adopted (e.g., QWERTY for English language countries, French AZERTY, German QWERTZ, and Italian QZERTY layouts – to reduce ambiguity we will, casually, refer to this family of keyboards as QWERTY keyboards). When well designed, small QWERTY keyboards can make text entry fast by giving the users good physical targets and feedback. However, there is a strong design trade-off between keys being large enough for fast, easy typing and overall device size, with large-fingered users often finding the keys simply too small to tap individually at speed. Physical keyboards also interact poorly with touch screens, where one hand often needs to hold a stylus, and they reduce the space available on the device for the screen.

The QWERTY keyboard layout was designed as a compromise between speed and physical characteristics of traditional manual typewriters: the layout separates commonly occurring pairs of letters to avoid head clashes on manual typewriters, and is imbalanced between left and right hands. Clearly, head clashes and manual carriage returns are not an issue with desktops nor handhelds, but their history in the design leads to a suboptimal layout, where users have to move their fingers more often than they would on an ideal layout. Faster keyboards have been designed, the most widely known being the Dvorak keyboard (or American Simplified Keyboard) for touch typists of English-language documents (Figure 1). While significantly faster than QWERTY keyboards, these have not been widely adopted, primarily because of the learning time and invested skill set in QWERTY keyboards. This investment has been shown to carry over into smaller devices, where the suboptimality issue is even stronger, as users tend to type with one or two thumbs, not the nine fingers envisaged of touch typists. There is strong evidence that alphabetic layouts on desktop computers give no benefits even for novice users, and hinder people with any exposure to the QWERTY layout (Norman, 2002; Norman & Fisher, 1982). It can be reasonably assumed this is also true of palmtops, although some research has shown that experience of using a desktop QWERTY keyboard gives no benefit when moved to a very new environment (McCaul & Sutherland, 2004). While optimal layouts could be designed around two-thumb entry, these are likely to be so different from users' experiences that initial use would be very slow and, as with the Dvorak, rejected by end users. Furthermore, these would

Figure 1. Dvorak keyboard



still be suboptimal for one-thumb use (or vice-versa if designed for one thumb).

One approach to small keyboard design that builds on QWERTY skills, and the imbalance between left and right hands on the QWERTY pad, is the half-QWERTY keyboard (Matias, MacKenzie, & Buxton, 1996). Here the keyboard is physically halved down the centre with only the left set of the letters given as physical keys, the user holds the space bar to *flip* the keyboard to give the right side letters (Figure 2-left), and permits fast one-handed entry. In experiments, users of the half-QWERTY keyboard were shown to quickly achieve consistent speeds of 30 words per minute or higher (when using a keyboard with desktop-sized keys). The FrogPad™ is a variant, using an optimised keyboard, so that use of the "right side" of the keyboard is minimised (Figure 2-right). Matias et al. (1996) predicted an optimised pad would lead to a speed increase of around 18% over the half-qwerty design, but at a cost of lost transferable skills. FrogPad™ Inc. now manufactures an optimised keyboard along these lines and claims 40+ words–per-minute typing speeds. Neither approach has yet to make it onto handheld devices, but the FrogPad™ is marketed as a separate keyboard for PDAs.

The FastTap™ keyboard, however, has been targeted at mobile devices from initial conception. This patented technology takes a different approach to miniaturisation by including an alphabetic keyboard as raised keys between the standard numeric keys of a phone pad, giving direct nonambiguous text entry on a very small platform while preserving the standard 12-key keypad currently used by over 90% of mobile users globally (see Figure 3. Experiments (Cockburn & Siresena, 2003) have shown that FastTap™ is considerably faster and easier to use for novice users than more standard predictive text approaches (discussed next), and the two approaches perform similarly for expert users (once practiced, FastTap users, in their trial, achieved 9.3 words per minute (wpm) with T9™ users achieving 10.8wpm, somewhat slower than in other trials, see next section for discussion of T9).

A drastically different unambiguous keyboard approach is to use chords, multiple simultaneous key presses mapping to a single character, either using one or both hands. One-handed chord

Figure 2. Simplified half-QWERTY and FrogPad™

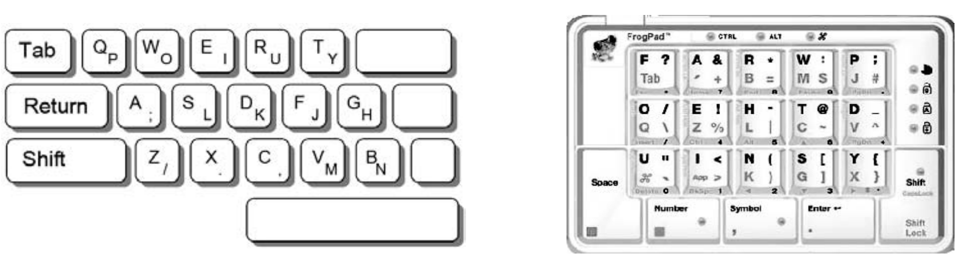


Figure 3. FastTap™ phone keyboard

keyboards were originally envisaged as the ideal partner to the mouse (Engelbart & English, 1968), allowing users to enter text and point at the same time. Chord keyboards can give extremely fast entry rates, with court stenographers reaching around 225 words per minute using a two-handed chord keyboard (compared to skilled QWERTY touch-typing speeds of around 135wpm and handwriting of about 15wpm). Single-handed chord keyboards are, by definition, palm sized, as users have to be able to have one digit on each key, and have been used in mobile devices (Figure 4 right, shows the AgendA organiser including an alphabetic keyboard surrounded by a chord keyboard). However, the learning time is prohibitive with few users willing to learn the chords required to use these keyboards. Furthermore, the keyboards are not usable without training, users cannot guess how to use them when first picking up a device. Thus, despite size and speed advantages, chord keyboards are generally considered too alien for mainstream devices, and rarely appear on consumer products.

Ambiguous Keyboards

The most common ambiguous keyboard, and the dominant keyboard for mobile phones, is still the telephony ISO/IEC standard 12-key phone keypad (e.g., Figure 5 left). Originally envisaged for name-based dialling of telephone area codes, not for text entry, this keyboard groups three or four letters on each of the physical keys *2* through *9* (with the *1, *, #,* and *0* keys typically acting as space, shift, and other control keys on mobile phones, depending on handset). The method of disambiguating the multiple letters per key is discussed later. However, to further complicate text entry, numbers are typically entered either in a special number-only mode or by pressing and holding the appropriate key. Recently some phones have been released with a slightly stretched mobile phone pad, typically with two extra columns, to give at most two letters per key plus extra space for control keys (Figure 5 right).

While the 12-key mobile phone pad is the smallest commonly found keyboard layout, there has been a history of research into very small

Figure 4. Sample chord keyboards (Douglas Engelbart and Microwriter AgendA)

Figure 5. 12-key phone pad (Nokia N73) and stretched phone pad (Blackberry™ 7100)

keyboards with as few as three keys. Text entry on these usually involves cursor movement through the alphabet. The most common approach is the 3-key date-stamp method widely used in video games (left and right scroll through an alphabetic strip of letters with fire entering the current letter) and 5-key variant using a joystick with a 2-D keyboard display. Experiments have been conducted to compare simple alphabetic layouts with layouts that are optimised for text entry in English, and ones that dynamically adjust to the most likely next letter (Bellman & MacKenzie, 1998; MacKenzie, 2002b). These studies showed that, once practiced, users of a 1-D alphabetic strip achieved around 9 words per minute while 2-D QWERTY users reached around 10-15 words per minute. Unfortunately, dynamic adjustment of the layout based on probabilities of next letter did not have the expected speed up in either 1-D or 2-D, due to attention load from the user slightly dominating the reduced time to select a letter. An alternative approach is to use short codes representing the letters, for example, short sequences of cursor keys. Evreinova, Evreino, and Raisamo (2004) showed that users could achieve good entry speeds with 3-key combinations of cursor keys, for example, left-up-left for *A*, and that, despite high initial error rates, users could learn the codes quickly.

Disambiguation

The traditional approach to disambiguating text entry on a mobile phone keypad is the manual multitap approach: users press keys repeatedly to achieve the letter they want, for example, on a standard phone keypad, *2* translates to *A* with *22* translating to *B*, and so forth. This approach has also been adopted in many other domestic devices such as audio/video remote controls. Multitap leads to more keystrokes, as users have to repeatedly click for most letters, and to a problem with disambiguating a sequence of letters on the same key, for example, *CAB* is *222222*. Users typically manually disambiguate this by either waiting for a timeout between subsequent letters on the same key or hitting a *time-out kill* button (say *); clearly an error-prone process and one that slows users down. Wigdor and Balakrishnan (2004) refer to multitap as an example of *consecutive* disambiguation; the user effectively enters a key then disambiguates it. An alternative manual disambiguation approach is *concurrent* disambiguation; here users use an alternative input method, for example, tilting the phone (Wigdor & Balakrishnan, 2003) or a small chord-keyboard on the rear of the phone (Wigdor & Balakrishnan, 2004), to disambiguate the letter as it is entered. While clearly potentially much faster than multitap and relatively easy to use, this approach has not yet been picked up by device manufacturers.

Aimed at overcoming the problems of multitap, predictive text-entry approaches use language modelling to map from ambiguous codes to words so that users need only press each key once, for example, mapping the key sequence *4663* directly to *good*. While there are clearly cases where there are more than one match to the numeric key sequence (e.g., *4663* also maps to *home* and *gone*, amongst others), these are surprisingly rare for common words. The problem of multiple matches can be alleviated to a large extent by giving the most likely word as the first suggestion, then allowing users to scroll through alternatives for less likely words. Based on a simple dictionary of words and their frequency of use in the language, users get the right word suggested first around 95% of the time (Gong & Tarasewich, 2005). AOL-Tegic's T9 (Grover, King, & Kushler, 1998; Kushler, 1998) industry-standard entry method is based around this approach and is now deployed on over two billion handsets. Controlled experiments have shown this form of text entry considerably outperforms multitap (Dunlop & Crossan, 2000; James & Reischel, 2001), with text entry rising from around 8 words per minute to around 20 for T9. While predictive text entry is very high quality, it is not perfect, and can lead to errors that are undetected, as users tend to type without monitoring the screen (e.g., a classic T9 user error is sending the message *call me when you are good* rather than *are home*). The main problem, however, with any word prediction system is handling out-of-vocabulary words: words that are not known to the dictionary cannot be entered using this form of text entry. The usual solution is to force users into an "add word" dialogue, where the new word is entered in a special window using multitap, clearly at a considerable loss of flow to their interaction and reduction in entry speed. As most people do not frequently enter new words or place/people names, this is not a major long-term problem. However, it does considerably impact on initial use and can put users off predictive text messaging, as they constantly have to teach new words to the dictionary in the early days of using a new device. This in turn impacts on consumer adoption, with many people not using predictive text, despite it clearly being faster for experienced users.

An alternative approach to dictionary and word-level disambiguation is to use letter-by-letter disambiguation, where letters are suggested based on their likelihood, given letters already entered in the given word or likely letters at the start of a word (e.g., in the clearest case in English, a *q* is most likely to be followed by a *u*). This gives the user freedom to enter words that are not in the dictionary, and considerably reduces the memory load of the text-entry system (no longer an issue with phones but still an issue on some devices). Experiments using this approach (MacKenzie, Kober, Smith, Jones, & Skepner, 2001) showed keystrokes halved and speed increased by around 36% compared with multitap. They also claim that this speed is inline with T9 entry, and that their approach outperforms T9 by around 30% when as few as 15% of the least common words are missing from the predictive dictionary. Predicting letters based on previous letters is actually a specific implementation of Shannon's approach to prediction based on n-grams of letters (Shannon, 1951): predicting the next letter based on the previous letters (or next word based on previous words). Some work has been carried out to extend this to the word level and shows good promise: for example bigram word prediction in Swedish with word completion reduced keystrokes by between 7% and 13% when compared with T9 (Hasselgren, Montnemery, Nugues, & Svensson, 2003).

Some work has been conducted on very small ambiguous keyboards following the same approach as predictive text on 12-key keypads. In work on watch-top text entry we (Dunlop, 2004) found that moving to a 5-key pad reduced accuracy from around 96% to around 81%, with approximately 40% reduction in text-entry speed. While this is a considerable drop in speed, this is a reasonable speed on such a small device, and it is considerably faster than picking from a virtual keyboard with a 5-way keypad (Bellman & MacKenzie, 1998). An interesting alternative input method for small devices is to use a touch-wheel interface, such as those on iPods™. Proschowsky, Schultz, and Jacobsen (2006) developed a method where users are presented with the alphabet in a circle, with a predictive algorithm hidden from the user. This algorithm increases the target area for letters based on the probability of them being selected next, so that users are more likely to hit the correct target when tapping on the touch wheel. User trials show around six to seven words per minute entry rates for novices, about 30% faster than the same users using a date-stamp approach on a touch wheel.

As with the QWERTY layout, the letters on an ambiguous phone keypad do not need to be laid out alphabetically. Here, however, the disambiguation method introduces an additional aspect to designing an optimal layout: the letters can be rearranged to minimise the level of ambiguity for a given language, in addition to looking at minimising finger movement. However, experiments predict that a fully optimised 8-key layout[1] would improve text-entry rates by only around 3% for English (Gong & Tarasewich, 2005). We found a larger but still small reduction of around 8% in keystrokes for a pseudo-optimised layout with the alphabet on four keys when compared against alphabetic ordering (Dunlop, 2004). This modelling does, however, show that stretching the standard phone pad from 8 to 12 keys for text entry is likely to result in a decrease of between 3% and 13% of keystrokes, depending on the style of language used (e.g., Figure 5 right) (Gong & Tarasewich, 2005).

STYLUS-BASED TEXT ENTRY

Compared to mobile phones, personal organisers (PDAs) have made more use of touch screens and stylus interaction as the basis of interaction, and this is now emerging on high-end phones such as Apple's iPhone. This frees up most of the device for the screen and leads to natural mouse-like interaction with applications. Lack of a physical keyboard has led to many different approaches for stylus-based text entry on touch-sensitive screens, which will be discussed in this section: on-screen (or soft) keyboards, handwriting recognition and more dynamic gesture-based approaches.

On-Screen Keyboards

A simple solution to text entry on touch screens is to present the user with an on-screen keyboard that the user can tap on with their stylus, or on particularly large touch screens with their finger. The most common implementation is to copy the QWERTY layout onto a small touch-sensitive area at the bottom of a touch screen. However, as with physical keyboards and keypads, there has been research into better arrangements of the keys. Mackenzie and his team have conducted a series of experiments on alternative on-screen keyboard layouts that are optimised for entry using a single stylus while one hand is holding the device (as opposed to desktop assumption of eight fingers and a thumb being used for entry). They investigated both unambiguous keyboards and an optimised 12-key ambiguous keypad, inspired by the success of T9™ and the fundamental rule of interaction that large targets are quicker to hit than small ones (Fitts, 1954). Their results estimate that an expert user could achieve 40+ wpm on a soft QWERTY keyboard with novice soft-keyboard users achieving around 20 wpm (MacKenzie, Zhang, & Soukoreff, 1999). The alternative layouts were predicted to give higher entry rates for expert use: the unambiguous Fitaly layout was predicted to reach up to 56wpm, and ambiguous JustType, 44wpm (Figure 6). However, novice users achieved only around 8wpm using these alternative keyboard layouts, highlighting the carry-over effect of desktop QWERTY layout.

While simple and fast, the on-screen keyboard approach can be tiring for users, as they are required to repeatedly hit very small areas of the screen. The patented technology underlying the

Figure 6. Fitaly and JustType keyboard layouts

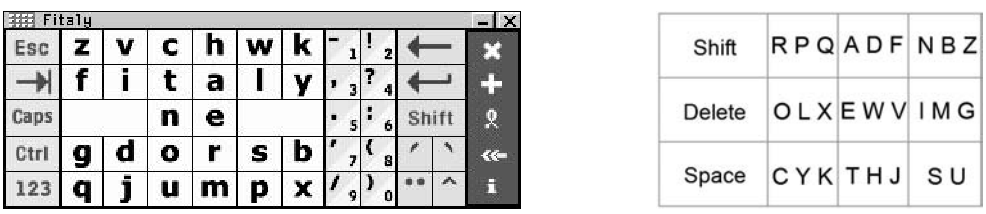
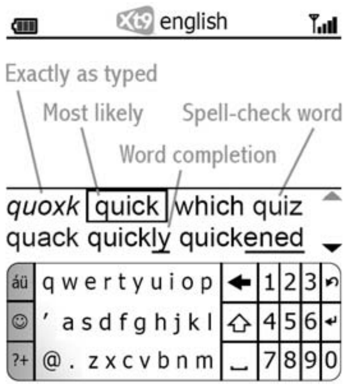


Figure 7. Sample XT9™ Mobile Interface



XT9™ Mobile Interface from Tegic Communications attempts to address this problem by including a level of disambiguation in an otherwise unambiguous keyboard (Robinson & Longe, 2000). For example, if the user taps a letter adjacent to the letter in the intended word, then the "intended" letter is used instead of the letter tapped. Their approach defaults to the most likely full word given the approximate letters entered, while offering alternative corrections and word completions as well as the letters actually typed (Figure 7). XT9 technologies have been developed by Tegic for multiple platforms, including handprinting and small physical keyboards.

Handwriting

To many, the obvious solution to text entry on handheld devices is handwriting recognition. However, there are many problems with basing text entry around handwriting,: most obvious being the slow rate at which people write (about 15 wpm (Card, Moran, & Newell, 1983)) and the wide variability of individual's handwriting styles. Modern handwriting recognition systems, for example on Windows XP Tablet edition, are extremely good at recognising in-dictionary words, but struggle on words that are not previously known, and are still inherently limited by writing speeds. Cursive handwriting recognition also requires a reasonably large physical space for the user to write on (governed by both touch-screen resolutions and human dexterity). Furthermore, cursive handwriting recognition still requires considerable processing power that is more in line with modern laptops/tablets than phones. Simplified alphabets to reduce the processing complexity and space needed for writing were originally introduced with the unistroke (Goldberg & Richardson, 1993) approach. Here each letter is represented as a single stroke, with letters typically drawn on top of each other in a one letter wide slot. This approach forces users to learn a new alphabet (Figure 8) but makes recognition computationally easier and more accurate, while also reducing the time it takes to draw each letter for skilled users. Palm popularised a more intuitive version, Graffiti™, on their palmtops: a mostly unistroke alphabet, Graffiti™, was composed mostly of strokes with high similarity to standard capital letters. CIC's Jot™ alphabet provides a mix of unistroke and multistroke letters and is deployed on a wide range of handhelds. Experiments comparing handprinting with other text-entry methods are rare, but a comparison between handprinting, QWERTY-tapping, and ABC-tapping on pen-based devices (MacKenzie, Nonnecke, McQueen, Riddersma, & Meltz, 1994) showed that a standard QWERTY layout can achieve around 23wpm while hand-

Figure 8. Unistroke, Graffiti™ and Jot™ sample letters

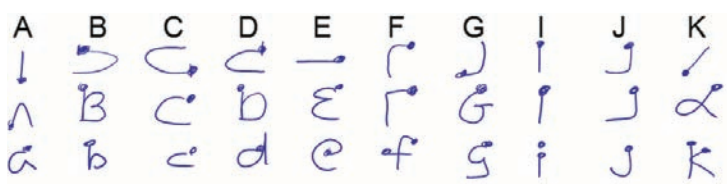
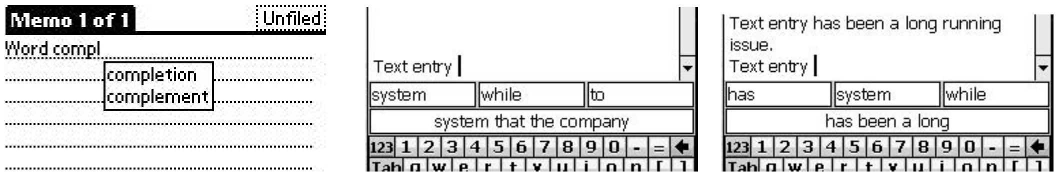


Figure 9. WordComplete™ (left) and AdapTex™ (centre and right)



printing achieved only 17wpm and alphabetic soft-keyboard only 13wpm.

Word completion can also be used to help users with text entry, and sits most comfortably with touch screens where users can point quickly at word completion suggestions. For example, CIC's WordComplete™ (Figure 9 left) suggests short phrases or word completions based on what the user has written. Similar technologies are used on the eZiType™[2] and XT9[3] technologies deployed on some mobile phones. While tempting, word completion needs to perform very well in order to give users a benefit: users rapidly get into a flow of text entry, which word completion interrupts, so the saving in terms of letters entered has to be considerable before a real saving in time is achieved. We (Dunlop & Crossan, 2000) estimated that simple word completion would reduce keystrokes by 17%, but our model-based evaluation (see section 4.1) predicted an approximate halving of entry speed once user interruption time was taken into account. Some recent advances, however, have shown that when based on more complex language models, word completion can be beneficial with novice users increasing typing speed by around 35% when using a soft keyboard (Dunlop, Glen, Motaparti, & Patel, 2006) (Figure 9 centre right). AdapTex™ performance also increases over time by learning patterns of use in the user's language to tune suggestions to the individual user and his/her context of use.

Gesture-Based Input

Gesture-based interaction attempts to combine the best of visual keyboards with easy-to-remember stylus movements to gain faster and smoother, while still easy-to-learn, text entry. Building on our motor memory for paths, approaches such as Cirrin (Mankoff & Abowd, 1998), Quikwriting (Perlin, 1998) and Hex (Williamson & Murray-Smith, 2005) are based on the user following a path *through* the letters of the word being entered (Figure 10). For on-screen approaches, this achieves faster entry rates than single character printing with reduced stress and fatigue when writing. Furthermore, in the case of Hex, the approach can be used one-handed on devices with accelerometers/tilt sensors.

Gestures can be combined with more conventional soft keyboards so that users can choose to tap individual letters, improving pick-up-and-use usability, or to enter words in one gesture by following the path of the letters on the touch keyboard (experts can then enter the gestures anywhere on screen) (Zhai & Kristensson, 2003).

Dasher is a drastically different approach to text entry that attempts to exploit interactive displays more than traditional text entry approaches. In Dasher (Ward, Blackwell, & MacKay, 2000) (Figure 12), letters scroll towards the user and (s)he picks them by moving the stylus up and down as the letters pass the stylus. The speed of scrolling

Figure 10. Quikwriting (left) and the Hex entry for "was" (omitting letter display)(right)

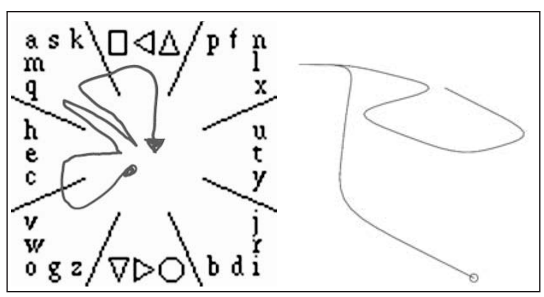


Figure 11. The ATOMIK keyboard with SHARK shortcuts

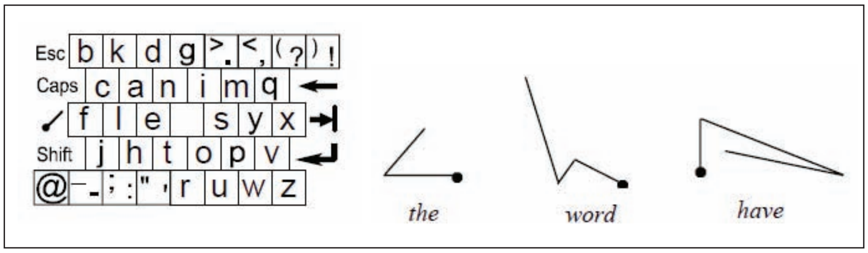
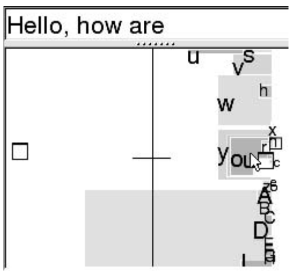


Figure 12. Dasher



is controlled by the user moving the stylus left and right with predictive text-entry approaches dynamically changing the space allocated to each letter (so that likely next letters are given more space than less likely ones, but all letters are available at each stage). Experiments show that users can enter at over 30 words per minute.

EVALUATION

Unlike many areas of mobile technology, where market forces and commercial ingenuity dominate, the field of text entry has benefited from considerable scientific study to establish the benefits of one method over another. These studies have been conducted by academic and industrial research groups, often in collaboration, and are used both to compare techniques and to tune their usage to how users actually enter text. Much of the related evaluation work and results have already been discussed previously; in this section, we focus on the evaluation methods themselves.

Technical Evaluation

The literature commonly uses three methods for reporting the performance of text entry: *average ranked list position* (ARP), *disambiguation accuracy* (DA), and *keystrokes per character* (KSPC).

The average ranked-list position (e.g., Dunlop & Crossan, 2000) for evaluating ambiguous text-entry methods is calculated in two phases. First language models, for example, in the simplest case, word frequencies, are learned from a corpus appropriate to the target language. Once trained, the second phase involves processing the same corpus one word at a time. Each word taken from the corpus is encoded using the ambiguous key coding for the target keypad (e.g., *home* is encoded as *4663*) and a ranked list of suggested words produced for that encoding based on the learned language model. The position of the target word in this list is averaged over all words to give the average ranked-list position for that corpus and keypad. An ARP value of 1.0 indicates that the correct word was always in the first position in the ranked list of suggestions, a value of 2.0 that, on average, the correct word was second in the ranked list. We predicted an ARP value of around 1.03 for a large corpus of English language newspaper articles using a standard phone keypad layout. ARP naturally biases the averaging process so that words are taken into account proportionally to their occurrence in the text corpus.

Disambiguation accuracy (e.g., Gong & Tarasewich, 2005) reports the percentage of times the first word suggested by the disambiguation process is the word the user intended: a DA value of 100% implies the disambiguation process always gives the correct word first, while 50% indicates that it only manages to give the correct word first half of the time. Gong and Tarasewich reported DA of 97% for written English corpus and 92% for SMS messages (both on a phone pad). This is a more intuitive and direct measure than ARP, but does not take into account the performance of words that do not come first in the list.

KSPC (MacKenzie, 2002a) reports the average number of keystrokes required to enter a character, for example, *home* followed by a space on a standard T9™ mobile phone requires six keystrokes – 4663*# where * is the next suggestion key and # space, giving a KSPC for *hello* of 6/4=1.5. As with ARP and DA, the value is normally averaged over a large corpus of appropriate text for the target language. A KSPC value of 1.0 indicates perfect disambiguation, as the user never needs to type any additional letters, while a higher figure reflects the proportional need for the next key in disambiguation (and a lower level, successful word completion). Full-sized nonambiguous keyboards achieve KSPC=1.00, standard date stamp method for entering text on three keys achieves KSPC=6.45, date stamp like interaction on five keys achieves KSPC=3.13 and multitap on a standard 9-key mobile phone achieves a KSPC of around 2.03 (MacKenzie, 2002a). Hasselgren et al. (2003) reported KSPC of 1.01 and 1.08 for T9 using Swedish news and SMS corpora respectively, improving to 1.01 and 0.88 respectively for their bigram model with word completion. KSPC does take into account ranked list position for all words, and compares easily with nonpredictive text-entry approaches; however, it is a rather abstract measure being based on letters, especially for dictionary-based approaches that are inherently word-based methods.

To gain an insight into potential expert user behaviour with different keyboards, different approaches have been taken to modelling interaction in order to predict expert (trained, error-free) performance. There are two basic approaches: physical movement modelling and keystroke level modelling. We (Dunlop & Crossan, 2000) proposed a keystroke level model based on Card, Moran and Newall's work (Card, Moran, & Newell, 1980). Our model was based on predicting the time *T(P)* taken by an expert user to enter a given phrase. The model calculates this in an equation that combines a set of small time measurements for elements of the user interaction. In the case of text entry, the relevant factors are the homing time for the user to settle on the keyboard T_h (0.40 seconds); the time it takes a user to press a key T_k (0.28s); the time it takes the user to mentally respond to a system action T_m (1.35s); the length of an average word k_w (4.98); and the number of words in the phrase w (10). In addition, for predictive text entry, where disambiguation occurs by the user moving through the ranked list of suggestions, the ARP value is required, here given as l=1.03. The overall time equation for entering a phrase is given in Equation 1 (Dunlop and Crossan's keystroke model):

$$T(P) = T_h + w\,(k_w T_k + l(T_m + T_k))$$

This model, as corrected by Pavlovych and Stuerzlinger (Pavlovych & Stuerzlinger, 2004), predicts a text-entry time for a 10-word phrase at 31.2 seconds, equating to a speed of 19.3 words per minute. This prediction matches closely with focused user experiments on experienced users of T9 of 20.4 wpm (James & Reischel, 2001).

We modelled keystroke speed at 0.28s based on a fixed figure from Card et al.'s work that is equivalent to "an average nonsecretary typist" on a full QWERTY keypad. This works well, however, it cannot take into account fine-grained keyboard design elements that can have a considerable impact on typing speed in practice: for example, different keyboard layouts clearly affect the average time it takes a user to move his/her fingers to the correct keys. Mackenzie's group have conducted considerable work using Fitt's law (Fitts, 1954) to calculate the limit of performance given distance between keys (e.g., Silfverberg, MacKenzie, & Korhonen, 2000). The basic form of their distance-based modelling predicts 40.6 wpm for thumb-based predictive input, assuming no *next* key operations (essentially equivalent to no thinking or homing times in equation 1). Later work modifies the Fitt's distance models to take into account two inaccuracies that can noticeably affect predictions: repeated letters on the same key (Soukoreff & MacKenzie, 2002), and parallel finger movements, where users move one finger at the same time as pressing with another (a main design criteria for QWERTY desktop keyboards that impacts on expert text users who use two thumbs) (MacKenzie & Soukoreff, 2002).

These models are useful in predicting performance but focus on expert error-free performance. More complex modelling approaches have been researched to support novices to model more complete interaction, and to model error behaviour (e.g., How & Kan, 2005; Pavlovych & Stuerzlinger, 2004; Sandnes, 2005). Although users studies are the traditional acid test for any interactive system, these models are valuable either in the early stages of design or to understand methods where user experiments are difficult (or biased by users' prior experience of current technologies).

User Studies

Models that predict text-entry performance only give us part of the picture, proper user studies often give a truer indication of how text-entry methods perform in reality. While there are many parameters that can affect the design of user studies, the three prominent ones for text-entry experiments are the environment in which the experiments are conducted, the platform on which experiments are conducted, and the source of the phrases that users enter.

Most user studies into text entry have been conducted in laboratories. A laboratory is a controlled environment that leads to a more consistent user experience and, thus, considerably easier statistical analysis as there are fewer confounding variables from the environment to interfere with measurements taken. However, conducting experiments on people entering text on mobile phones in quiet office settings where they can focus exclusively on the text-entry tasks is arguably not representative of normal use. There is a growing debate in mobile HCI research on the validity of laboratory experiments with some researchers arguing that, while the focus of most common errors is different in the real world, laboratory experiments do not miss errors that are found in real-world experiments (Kaikkonen, Kekäläinen, Cankar, Kallio, & Kankainen, 2005) while others claim a wider range of errors were found in the real world than in laboratories (Duh, Tan, & Chen, 2006). Kjeldskov and Graham (2003) report that "71% [of studied evaluations were] done through laboratory experiments, 19% through field experiments and the remaining 10% through surveys." As a specific example, Brewster (2002) showed usability and text-entry rates were significantly reduced for users performing an outdoor walking circuit, while entering on a soft numeric key pad, than those conducting the same experiment in a traditional laboratory. Whereas Mizobuchi, Chignell, and Newton (2005) showed that, while walking was slowed down when using a device, it did not impinge upon the text-entry rate or accuracy.

Experiments are either conducted on a real device or on a simulator; either running on a desktop PC or on a touch screen PDA. Obviously, experimenting on real devices is preferable in terms of appropriateness of experimental setting, however, development on real devices was very difficult until recently, as programming these by anyone but the manufacturer was difficult. However, most modern phones and PDAs have powerful Java platforms available, as standard, that can be used for controlled experiments. Conducting experiments on desktops leads to problems of users typing too fast: desktop keys are typically bigger and give a more positive response than many mobile phone keyboards, thus, potentially significantly increasing key speeds, which may bias some text-entry methods over others. There is also a problem with computer numeric pads being vertically inverted compared to phone pads; a historic difference that most users are unaware of that can become very visible when forced to use a computer keypad for phone operations. Emulating on a touch-screen handheld is a tempting alternative as users can hold the device naturally, and the interface designer has full freedom of key layout, however, without physical feedback of key presses and physically distinguished hit areas, typing can be considerably slower, again biasing some input methods over others.

Finally, users are typically required to enter a set of phrases on devices to measure their text-entry speed. These phrases are usually the same for all users in a trial so that variations in phrases can be excluded in statistical analysis. While there is no widespread agreement on phrases that are used, MacKenzie and Soukoreff (2003) proposed a standard set of short phrases that has been used by other researchers and provides a valuable baseline for comparisons. One problem with mobile phone text entry is that it is often used for short casual messages and testing with formal phrases from a traditional text corpus is not appropriate (see differences discussed previously for those who have experimented with formal English and SMS). This is compounded by the original multitap text-entry approach and short length of text messages[4] leading to considerable use of, often obscure, abbreviations that are not normally found in a corpus. To address this, How and Kan (2005) developed a large set of phrases extracted from SMS users' real text conversations. Although somewhat skewed to local Singaporean phrases and abbreviations (much of *SMS speak* is heavily localised and even personalised within a group of friends), the corpus is a valuable insight into the language often used on mobile phones. It should also be noted that entry speeds of 33wpm for users when transcribing text on desktop keyboards have been found to drop to around 19wpm for composing new text (Karat, Halverson, Horn, & Karat, 1999), so most results from text-entry experiments can be assumed to be over-inflating speeds by around 40% as they are typically based on transcription.

CONCLUSION AND FUTURE TRENDS

This chapter has reviewed a large number of text-entry methods that range from standard methods, through slight variations, to completely different approaches. We have looked at different hardware keyboard designs, different on-screen keyboard layouts, handwriting-based approaches, and more novel approaches such as gestures. We have also looked at ambiguous and unambiguous designs and the related approaches to disambiguation. Much of the work reported has experimental backing to show the potential benefits of each approach. However, when comparing the wide diversity of approaches in the literature to widely available implementations on real devices, the overriding message we see is that guessability, the initial pick-up-and-use usability of hardware/software, is paramount to success.

It is extremely hard to predict future trends for mobile devices: while there is considerable research showing the benefits and strengths of different approaches, market forces and the views of customers and their operators have a major role in deciding which techniques become widely adopted. Predicted gains in expert text-entry performance are of no use if people do not understand how to use the text-entry approach out of the box. To this end, we see considerable scope for entry methods that provide a smooth transition from novice to expert performance: XT9™ is one successful example of novice-to-expert support, as users get faster they will learn to be sloppier and type faster, without necessarily being consciously aware of why. Context-aware word completion that learns about individuals is another area that shows good potential: good for slow novice typists as they start, but building context and personalising as they gain proficiency.

Finally, looking at current market directions and the increasing desire to enter more text on small devices, we see the 12-key keypad slowly disappearing from phones, to be replaced with less number-centric entry methods. Despite its suboptimality and problems on small devices, both market trends and some user tests point to the QWERTY keyboard taking on this role, either as a physical or an on-screen keyboard.

PROJECTS AND DISCUSSION TOPICS

To build your experience of the material covered in this chapter, you could try the following small experiments:

1. Ask friends/colleagues/family how they normally enter text messages on their mobile phone, then time them entering a common phrase, say a nursery rhyme. Record times and error rates. Swap your users to another input method and see how their time and error rates compare. You can now compare expert vs. (semi-) novice performance on a range of techniques.
2. A variation of (1) would be to have a shoot-out, where you get a group together, agree a phrase, then shout "start" to get the whole group entering the phrase at once then raising their device when finished. You might want to control error policy by saying that all errors should be corrected before hands are raised. You might also want to include a word that is not in the dictionary, say an unusual local place name.
3. Search for some word completion software for your mobile device platform (e.g., Windows Mobile, PalmOS, or Symbian). Install this and work consistently, using it for 1 week. Record your views on the software after 10 minutes, 1 hour, 1 day, and 1 week. You can now compare your views as you learned the software (and it possibly learned your vocabulary/language).

You might also consider the following reflective comments/discussion topics (they are intentionally aggressive!):

1. T9™ is only for softies, real texters use multitap
2. QWERTY keypads on phones are included mainly so that business users do not look like teenagers sending text messages
3. Technical usability evaluation is pointless, only real user studies are valid for assessing different text-entry techniques
4. If the Fitaly keyboard is so much faster for expert users and is usable out of the box, then phone manufacturers should be bold and drop the QWERTY keyboard.

ACKNOWLEDGMENT

Our thanks are extended to our reviewers, whose comments have strengthened this chapter. We are also grateful to those organisations and companies that have given permission to use images throughout the chapter.

All trademarks are used with appropriate reference to their holders. FrogPad is a trademark of FrogPad Inc.—Figure 2 (right) is used with permission of FrogPad. FastTap is a trademark of Digit Wireless Inc. Figure 3 used with permission of Digit Wireless. Douglas Engelbart image (Figure 4 left) is used courtesy of SRI International, Menlo Park, CA. MicroWriter AgendA image (Figure 4 right) used with permission of Bellaire Electronics. Blackberry is a trademark of Research In Motion Limited. Fitaly keybord image (Figure 6 left) used with permission of Textware Solutions. XT9 is a registered trademark of Tegic Communications, Inc. The Tegic Communications, Inc. name, icons, trademarks, and image in Figure 7 are used with permission. Microsoft Windows is a registered trademark of Microsoft Corp. Unistroke is patented technology of Xerox. Graffiti is a registered trademark of Palm Inc; Jot and WordComplete of CIC. Figure 9 (centre and right) images used courtesy of KeyPoint Technologies, of which AdapTex is a trademark. Figure 10, 11, and 12 images used courtesy of related paper authors. Figures 1, 2 (left), 6 (right), and 8 drawn by authors and may be altered from original layouts for clarity. Figures 9 (left), 10 (left), and 12 screen images taken by authors. Images in Figure 5, original photographs by authors.

REFERENCES

Bellman, T., & MacKenzie, I. S. (1998). *A probabilistic character layout strategy for mobile text entry*. Paper presented at the Graphics Interface '98, Toronto, Canada.

Brewster, S. A. (2002). Overcoming the lack of screen space on mobile computers. *Personal and Ubiquitous Computing, 6*(3), 188-205.

Card, S. K., Moran, T. P., & Newell, A. (1980). The keystroke-level model for user performance time with interactive systems. *Communications of the ACM, 23*(7), 396-410.

Card, S., Moran, T., & Newell, A. (1983). *The psychology of human-computer interaction*. Lawrence Erlbaum Associates.

Cockburn, A., & Siresena, A. (2003). *Evaluating mobile text entry with the Fastap keypad*. Paper presented at the People and Computers XVII (Volume 2): British Computer Society Conference on Human Computer Interaction, Bath, England.

Duh, H. B., Tan, G. C., & Chen, V. H. (2006). *Usability evaluation for mobile device: A comparison of laboratory and field tests.* Paper presented at the 8th Conference on Human-Computer interaction with Mobile Devices and Services: MobileHCI '06, Helsinki, Finland.

Dunlop, M. D. (2004). *Watch-top text-entry: Can phone-style predictive text-entry work with only 5 buttons?* Paper presented at the 6th International Conference on Human Computer Interaction with Mobile Devices and Services: MobileHCI 04, Glasgow.

Dunlop, M. D., & Crossan, A. (2000). Predictive text entry methods for mobile phones. *Personal Technologies, 4*(2).

Dunlop, M. D., Glen, A., Motaparti, S., & Patel, S. (2006). *AdapTex: Contextually adaptive text entry for mobiles.* Paper presented at the 8th Conference on Human-Computer interaction with Mobile Devices and Services: MobileHCI '06, Helsinki, Finland.

Engelbart, D. C., & English, W. K. (1968). *A research center for augmenting human intellect.* Paper presented at the AFIPS 1968 Fall Joint Computer Conference, San Francisco, USA.

Evreinova, T., Evreino, G., & Raisamo, R. (2004). *Four-key text entry for physically challenged people*. Paper presented at the 8th ERCIM workshop: User Interfaces for All (Adjunct workshop proceedings), Vienna, Austria.

Fitts, P. M. (1954). The information capacity of the human motor system in controlling the amplitude of movement. *Journal of Experimental Psychology, 47*(6), 381-391.

Goldberg, D., & Richardson, C. (1993). *Touch-typing with a stylus.* Paper presented at the INTERCHI93.

Gong, J., & Tarasewich, P. (2005). *Alphabetically constrained keypad designs for text entry on mobile devices.* Paper presented at the CHI '05: SIGCHI Conference on Human Factors in Computing Systems, Portland, USA.

Grover, D. L., King, M. T., & Kushler, C. A. (1998). *USA Patent No. 5818437. I. Tegic Communications.*

Hasselgren, J., Montnemery, E., Nugues, P., & Svensson, M. (2003). *HMS: A predictive text entry method using bigrams.* Paper presented at the Workshop on Language Modeling for Text Entry Methods, 10th Conference of the European Chapter of the Association of Computational Linguistics, Budapest, Hungary.

How, Y., & Kan, M.-Y. (2005). *Optimizing predictive text entry for short message service on mobile phones.* Paper presented at the Human Computer Interfaces International (HCII 05), Las Vegas, USA.

James, C. L., & Reischel, K. M. (2001). *Text input for mobile devices: Comparing model prediction to actual performance*. Paper presented at the SIGCHI conference on Human factors in computing systems, Seattle, Washington, USA.

Kaikkonen, A., Kekäläinen, A., Cankar, M., Kallio, T., & Kankainen, A. (2005). Usability testing of mobile applications: A comparison between laboratory and field testing. *Journal of Usability Studies, 1*(1), 4-17.

Karat, C., Halverson, C., Horn, D., & Karat, J. (1999). *Patterns of entry and correction in large vocabulary continuous speech recognition systems.* Paper presented at the ACM CHI99: Human Factors in Computing Systems.

Kjeldskov, J., & Graham, C. (2003). *A review of mobile HCI research methods.* Paper presented at the Fifth International Symposium on Human Computer Interaction with Mobile Devices and Services: MobileHCI 03, Udine, Italy.

Kushler, C. (1998). *AAC using a reduced keyboard.* Paper presented at the Technology and Persons with Disabilities Conference, USA.

MacKenzie, I. S. (2002a). *KSPC (keystrokes per character) as a characteristic of text entry techniques.* Paper presented at the MobileHCI 02: Fourth International Symposium on Human-Computer Interaction with Mobile Devices, Pisa, Italy.

MacKenzie, I. S. (2002b). *Mobile text entry using three keys.* Paper presented at the Proceedings of the second Nordic conference on Human-computer interaction, Aarhus, Denmark.

MacKenzie, I. S., Kober, H., Smith, D., Jones, T., & Skepner, E. (2001). *LetterWise: Prefix-based disambiguation for mobile text input.* Paper presented at the 14th annual ACM symposium on User interface software and technology Orlando, USA.

MacKenzie, I. S., Nonnecke, R. B., McQueen, C., Riddersma, S., & Meltz, M. (1994). *A comparison of three methods of character entry on pen-based computers.* Paper presented at the Factors and Ergonomics Society 38th Annual Meeting, Santa Monica, USA.

MacKenzie, I. S., & Soukoreff, R. W. (2002). *A model of two-thumb text entry.* Paper presented at the Graphics Interface 2002.

MacKenzie, I. S., & Soukoreff, R. W. (2003). *Phrase sets for evaluating text entry techniques.* Paper presented at the ACM Conference on Human Factors in Computing Systems–CHI 2003, New York.

MacKenzie, I. S., Zhang, S. X., & Soukoreff, R. W. (1999). Text entry using soft keyboards. *Behaviour & Information Technology, 18,* 235-244.

Mankoff, J., & Abowd, G. D. (1998). *Cirrin: A word-level unistroke keyboard for pen input.* Paper presented at the 11th Annual ACM Symposium on User interface Software and Technology: UIST '98, San Francisco.

Matias, E., MacKenzie, I. S., & Buxton, W. (1996). One-handed touch typing on a QWERTY keyboard. *Human-Computer Interaction, 11*(1), 1-27.

McCaul, B., & Sutherland, A. (2004). *Predictive text entry in immersive environments.* Paper presented at the IEEE Virtual Reality Conference 2004 (VR 2004), Chicago, USA.

Mizobuchi, S., Chignell, M., & Newton, D. (2005). *Mobile text entry: Relationship between walking speed and text input task difficulty.* Paper presented at the MobileHCI 05: 7th international Conference on Human Computer interaction with Mobile Devices and Services Salzburg, Austria.

Norman, D. A. (2002). *The design of everyday things.* Basic Books.

Norman, D. A., & Fisher, D. (1982). Why alphabetic keyboards are not easy to use: Keyboard layout doesn't much matter. *The Journal of the Human Factors Society, 25*(5), 509-519.

Pavlovych, A., & Stuerzlinger, W. (2004). *Model for non-expert text entry speed on 12-button phone keypads.* Paper presented at the CHI '04: SIGCHI Conference on Human Factors in Computing Systems, Vienna, Austria.

Perlin, K. (1998). *Quikwriting: Continuous stylus-based text entry.* Paper presented at the 11th Annual ACM Symposium on User interface Software and Technology: UIST '98, San Francisco.

Proschowsky, M., Schultz, N., & Jacobsen, N. E. (2006). *An intuitive text input method for touch wheels.* Paper presented at the SIGCHI Conference on Human Factors in Computing Systems: CHI '06, Montréal, Canada.

Robinson, B. A., & Longe, M. R. (2000). *USA Patent No. 6,801,190/7,088,345.*

Sandnes, F. E. (2005). *Evaluating mobile text entry strategies with finite state automata.* Paper presented at the MobileHCI 05: 7th international Conference on Human Computer interaction with Mobile Devices and Services, Salzburg, Austria.

Shannon, C. (1951). Prediction and entropy of printed english. *Bell System Technical Journal*, *30*, 50-64.

Silfverberg, M., MacKenzie, I. S., & Korhonen, P. (2000). *Predicting text entry speed on mobile phones.* Paper presented at the SIGCHI Conference on Human Factors in Computing Systems (CHI'00), The Hague, The Netherlands.

Soukoreff, R. W., & MacKenzie, I. S. (2002). *Using Fitts' law to model key repeat time in text entry models.* Paper presented at the Graphics Interface 2002.

Ward, D. J., Blackwell, A. F., & MacKay, D. J. (2000). *Dasher—A data entry interface using continuous gestures and language models.* Paper presented at the 13th Annual ACM Symposium on User interface Software and Technology: UIST '00, San Diego, USA.

Wigdor, D., & Balakrishnan, R. (2003). *TiltText: Using tilt for text input to mobile phones.* Paper presented at the 16th annual ACM Symposium on User Interface Software and Technology (UIST '03), New York.

Wigdor, D., & Balakrishnan, R. A. (2004). *Comparison of consecutive and concurrent input text entry techniques for mobile phones.* Paper presented at the SIGCHI Conference on Human Factors in Computing Systems (CHI '04), New York.

Williamson, J., & Murray-Smith, R. (2005). *Dynamics and probabilistic text entry.* Paper presented at the Hamilton Summer School on Switching and Learning in Feedback systems.

Zhai, S., & Kristensson, P. (2003). *Shorthand writing on stylus keyboard.* Paper presented at the SIGCHI Conference on Human Factors in Computing Systems: CHI '03, Ft. Lauderdale, Florida, USA.

KEY TERMS

Ambiguous Keyboards: A keyboard layout where each key is related to many letters (e.g., the standard 12-key phone pad layout where, say, 2 is mapped to ABC)

Evaluation: Method for assessing and measuring the performance of a text-entry system in terms of either usability or technical performance.

Handwriting Recognition: Method for interpreting text that has been entered using handwriting via stylus

Predictive Text Entry: Text-entry method that attempts to predict the user's intended words from an ambiguous input (sometimes extended to predict word or phrase completions).

Text Entry: Method of inputting text to a mobile device

Unambiguous Keyboards: A keyboard layout where one key-press unambiguously relates to one character (e.g., 2 is mapped to A)

Usability Techniques: Series of methods and tools for designing and evaluating the usefulness and effectiveness of a text-entry system.

User Studies: Evaluations that are conducted to assess the performance of a system with real end users, generally conducted in usability laboratories under controlled settings.

ENDNOTES

[1] The standard 12-key phone key pad uses 8 keys for letters

[2] http://www.zicorp.com/eZiType.htm

[3] http://www.tegic.com/products/xt9.asp

[4] Original SMS (Short Message Service) messages were limited to 160 characters.

Chapter XXVI
Improving Stroke-Based Input of Chinese Characters

Min Lin
University of Maryland, Baltimore County (UMBC), USA

Andrew Sears
University of Maryland, Baltimore County (UMBC), USA

Steven Herbst
Motorola, USA

Yanfang Liu
Motorola China Electronics Ltd., PR China

ABSTRACT

This chapter presents a case study of the redesign of the mobile phone keypad graphics that support the Motorola iTap™ stroke-based Chinese input solution. Six studies were conducted to address problem identification, proof of concept evaluation, usability testing in both US and China, and design simplification to support business objectives. Study results confirmed that a new abstract-with-examples design helped users to develop more accurate knowledge regarding stroke-to-key mappings and lead to significant improvements in both text-entry speed and accuracy. The data also showed that, when using the new keypad graphics, the stroke-based input method could outperform the popular Pinyin technique after about 1 hour of casual usage, making the stroke method a competitive alternative for Chinese entry on mobile phones.

INTRODUCTION

The capabilities of modern mobile phones are far beyond the literal combination of "mobile" and "phone." Rapid improvements in both hardware and software have turned mobile phones into personal multichannel communication centers. Mobile phones now serve many roles including music player, camera, camcorder, voice recorder, game player, calendar, notepad for short notes, and text, e-mail, and IM messaging device. As a result, the global mobile phone market is growing rapidly. According to iSuppli, over 800 million mobile phones were shipped in 2005, a 14% increase compared to 2004. China is the largest mobile phone market in the world. Due to the growing popularity of short message service (SMS), increased support for personal information management (PIM), and Internet browsing capabilities, effective text-entry techniques are becoming more and more important. Researchers have investigated a variety of alternatives for entering English text using the limited number of keys available on mobile phones, confirming that existing solutions can be awkward and slow (James and Reischel, 2001; MacKenzie, Kober, Smith, Johns, & Skepner, 2001; Silfverberg, MacKenzie, & Korhonen, 2000). While entering English text can be challenging, entering Chinese characters using mobile phone keypad is much more difficult. Instead of a relatively modest 26 letters plus numbers and a few symbols, Chinese entry requires the user to learn how to enter thousands of characters, a task for which professionals may use a keyboard with as many as 4,000 keys (Archer, Chan, Huang, & Liu, 1988).

In this chapter, we present the redesign of the keypad graphics (the symbols printed on the keys as the legends for Chinese strokes) for the Motorola iTap™ stroke-based input solution as a case study. This chapter expands significantly on an abbreviated version published earlier (Lin & Sears, 2005b). This chapter includes more detail and covers the entire process, from problem identification to solution development, proof of concept evaluation, and a series of user studies. This collaborative effort involved UMBC faculty and students, as well as Motorola employees in both the US and China.

The underlying problem was defined collaboratively, with input from Motorola and UMBC personnel. Motorola provided financial support, allowing UMBC faculty and students to conduct the initial studies that lead to the design of the new keypad graphics. Once the efficacy of the graphics was confirmed through US-based studies, Motorola provided resources and personnel to replicate the UMBC studies in Beijing, China. Raw data from the Beijing studies was sent to UMBC for analysis. UMBC and Motorola personnel subsequently collaborated to develop presentations describing the results of these studies, which ultimately resulted in the new keypad graphics being formally adopted as the preferred solution for future Motorola mobile phones for the Chinese market. These graphics have already been used in several new phones.

This chapter starts by presenting a brief introduction to Chinese characters and various text techniques for entering Chinese text on mobile platforms. Next, a case study is presented that describes the redesign of the keypad graphics that support Motorola's iTap™ software. The case study involved (1) a comparison of the original iTap™ solution and the popular Pinyin method; (2) a test of an alternative design that was developed based on observations from the initial study; (3) a 6-day, longitudinal hands-on study of the proposed design; (4) a duplication of the longitudinal test that was conducted in China; (5) a study designed to simplify the proposed design to allow it to fit on smaller keypads; and (6) a final evaluation of the simplified design. For each study, we present the experimental design, data analysis, and a discussion of the results. We conclude by summarizing the experience presented in this case study.

CHINESE CHARACTERS

Chinese differs significantly from western languages such as English. Chinese is an ideographic language, with the shape of each character playing a critical role in presenting the meaning of the

Figure 1. Examples of the sound, shape, and meaning for Chinese characters

抱　饱

character. The shape often determines pronunciation as well. For example, the two characters on figure 1 have the same component on the right, differing only on the left. These two characters mean "hold in arms" and "well fed" (left and right characters, respectively), matching the meanings of the differing components on the left side of the characters ("hand" and "food", respectively). The shared right component is pronounced "bao," while both characters are pronounced as "bao" with different tones.

In Chinese, a stroke is the minimum structural unit and a character is the minimum functional unit. While there are thousands of Chinese characters, there are only 36 unique strokes. Most characters are composed of two or more strokes. For example, the two characters in figure 1 are both written using eight strokes. In addition, for any given character, there is a predefined "correct" order in which the strokes should be written. Therefore, if thinking of strokes as letters in English, then writing characters using strokes is conceptually the same as writing words using letters.

MOBILE INPUT SOLUTIONS

Handwriting Recognition

For mobile phones with a touch screen, such as smart phones, handwriting recognition is widely used for input, including tools like DragonPen™ and PenPower for Chinese input. With incremental recognition techniques, these tools may recognize a character without writing all the strokes (Matić, Platt, & Wang, 2002). However, like speech, the technology is unlikely to provide error-free input. More importantly, handwriting requires both hands: one to hold the device and the other to write. As a result, when one hand is occupied, handwriting can be difficult or impossible to use for effective input. In addition to usability, economic concerns may limit the use of handwriting recognition, since touch-sensitive screens can increase the cost of producing mobile phones.

Pinyin Method

Due to the nature of the Chinese language, there is no natural mapping between Chinese characters and the letters printed on the standard QWERTY keyboard or the standard telephone keypad. Therefore, the process of "Romanization" was used to define the pronunciation of Chinese characters using the letters of the Roman alphabet. Although some people considered such a process "peculiar" (Sacher, Tng, & Loudon, 2001), an official standard system, Pinyin, was created based on a northern dialect (i.e., Mandarin) and has been taught in primary schools in China for decades.

With Pinyin, every character can be entered using the letters of the standard mobile phone keypad. However, two major issues hinder performance when using Pinyin. Although the Mandarin pronunciation of each Chinese character is used to define Pinyin, it is not spoken by everyone in China. There are more than 50 dialects in China, and a large number of people either do not speak Mandarin or speak it in a nonstandard way. As a result, these people face significant difficulty in "translating" their pronunciations into Pinyin. For example, about 99% Hong Kong people did not speak Mandarin as their primary language (Census and Statistics Department, 2001). Even if this translation is not an issue, using Pinyin on a mobile phone is not as straightforward as entering English.

While the standard telephone keypad provides 12 keys, just 8 of these keys are used to represent the 26 letters of the Roman alphabet. As a result, each key represents several letters, resulting in ambiguity when a key is pressed. For English text entry, several input techniques have been developed to address this challenge and to speed up the process of entering English text. For example, Multitap allows a user to select one of several letters represented by a single key by pressing the key multiple times (Silfverberg et al., 2000). Less-Tap rearranges the letters on each key to allow for faster entry speeds (Pavlovych & Stuerzlinger, 2004). Predictive solutions, such as T9® by Tegic Communications, the iTap™ solution by Motorola, and eZiText® by Zi Corporation, automatically present a set of words that can be generated based on a sequence of key presses. Using these techniques, users press just one key per letter and an internal dictionary is used to determine which words the user might have been entering (James & Reischel, 2001). Like standard word completion applications, LetterWise guesses the next letter based on letters that have already been entered (MacKenzie et al., 2001). TiltText utilizes the orientation of the phone to resolve ambiguity. Users tilt the phone in one of the four directions to choose a specific letter following each key press (Wigdor & Balakrishman, 2003). Fastap™ from Digit Wireless uses more keys so each letter and number is represented by a unique key (Cockburn & Siresena, 2003). While results appear promising, adding more keys can become problematic as mobile phones become smaller.

By allowing users to enter English text more quickly, each of the techniques described could directly or indirectly speed up the process of entering Chinese text using Pinyin. Some techniques, such as TiltText and Fastap™, could provide direct benefits because they allow users to enter any random sequence of Roman characters more quickly (i.e., they are not based on a dictionary of English words). Other techniques, such as T9®, Less-Tap, and LetterWise are based on dictionaries and would need to be adapted to work with Pinyin scripts (the sequence of Roman letters used to represent the sound of a Chinese character) instead of English words. However, entering Pinyin is more complex than entering a simple sequence of Roman letters. First, the ambiguity introduced by multiple Roman letters sharing a key must be addressed. This is typically accomplished by choosing the desired Pinyin script from a set of possibilities associated with the sequence of keys that were pressed. Most Pinyin implementations facilitate this process using predictive capabilities similar to those provided by T9® and eZiText®. With these implementations, as users enter Roman letters, possible Pinyin scripts are presented, and the desired script can often be selected before the entire script has been entered. However, a single Pinyin script can still produce multiple Chinese characters. As a result, once the Pinyin script has been entered, the user must choose the correct character from a list of alternatives. The ambiguity introduced by multiple Roman letters sharing keys, combined with the fact that each Pinyin script may correspond to multiple Chinese characters, significantly increases the number of keystrokes required, and also forces users to shift their attention between the alternative lists and keys. Given the number of steps involved, mistakes can be hard to find and difficult to correct.

Structure-Based Methods

Unlike the many-to-many relationship between characters and pronunciations, the relationship between characters and structures is one-to-one. More importantly, the structure of every character is independent from its pronunciation, so structure-based solutions are not affected by the dialect that an individual speaks. Therefore, solutions based on the structure or shape of the character, instead of sound produced when it is spoken, can be used by people who do not speak Mandarin.

The primary structure-based method for normal QWERTY keyboard is Wubi (means “five keystrokes,” literally). Wubi uses some arbitrary rules to decompose characters into a set of substructures that are mapped onto the 26 Roman letters. Although expert users of Wubi can type much faster than those using Pinyin method, Wubi is very difficult to learn, as users have to memorize numerous, arbitrary, decomposition rules. Wubi is not practical for mobile phone users since they are not likely to spend multiple hours to learn an input method.

MOTOROLA ITAP™ STROKE INPUT METHOD

The Motorola iTap™ stroke input method (“stroke method” thereafter) enables users to enter Chinese characters stroke by stroke. For historical reasons, many Chinese characters can be written in two ways, using either simplified Chinese (official version of Chinese in the People’s Republic of China) or traditional Chinese (widely used elsewhere). While some characters remain the same in both versions, most differ significantly in appearance. The major difference is that simplified Chinese uses fewer strokes than traditional Chinese when writing the corresponding character (Figure 2), but both character sets are written using the same set of strokes. Therefore, any stroke-based input method can be used to write both simplified and traditional Chinese characters. While Pinyin can also be used to enter both sets of characters, people who use traditional Chinese typically use a different pronunciation annotation system, called Bopomofo, rather than Pinyin. Finally, the process of using stroke-based input may sound intuitive but a challenge exists, since the number of strokes is larger than the number of keys on a standard mobile phone keypad. As a result, multiple strokes must be assigned to each key.

In the stroke method, strokes are grouped and then assigned to the number keys, with a legend placed next to the number (Figure 3). Using predictive technology, a list of possible characters is presented and updated after each stroke input. Two keys (left and right arrow keys) are used to navigate through the list for character selection.

Figure 2. Three examples of corresponding characters in simplified Chinese and traditional Chinese

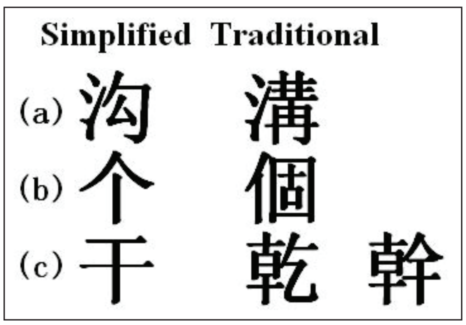
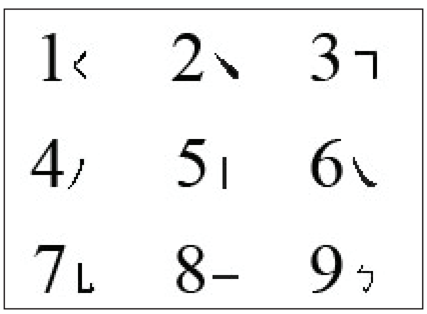


Figure 3. Motorola keypad design for iTap™stroke input method circa 2000



PRELIMINARY COMPARISON

In theory, structure-based methods should provide advantages over Pinyin, but no studies have been reported comparing data entry rates for the two solutions when used on a mobile phone. Therefore, our first study compared the stroke-based solution to Pinyin to provide a foundation for this project. While both Pinyin and stroke-based solutions can support the entry of both traditional and simplified Chinese characters, our study focuses on the use of simplified Chinese.

Participants

Thirty native Chinese speakers living in the Maryland area volunteered to participate in the study. Participants were randomly assigned to two groups. The first group completed the task

Table 1. The gender and age distribution of the two groups of participants

Input Method	Gender		Age Range (mean)
	Male	Female	
Pinyin	9	6	24-35 (27)
Stroke	5	10	24-31 (27)

using Pinyin, while the second used the stroke method (Table 1).

Task

The experimental task involved entering an e-mail message, comprised of 59 Chinese characters and 8 punctuation marks, using a mobile phone that supported both Pinyin and stroke method. Before starting the task, participants were given a brief verbal explanation of the keys used for navigation and character selection. The explanation did not address how characters or strokes were entered. Participants were given 10 minutes to practice using the phone. The e-mail message was presented on paper and was available throughout the task.

Results

The data revealed that participants using the Pinyin method were able to complete the task significantly faster than those using the stroke method [$t(28) = -4.1$, $p<0.001$]. The character entry speeds for Pinyin and the stroke method were 5.5cpm (characters per minute) and 2.7cpm, respectively. In addition, the error rate for the Pinyin group was significantly lower than that of the stroke method group [$t(28) = -3.0$, $p<0.01$]. The error rate of the Pinyin group was 1.5% while the error rate for the stroke group was 7.6%. The slow and error prone results of the participants using the stroke method raised significant concerns and motivated an in-depth investigation, with the goal of determining the factors that lead to such poor performance.

Since Pinyin is taught in primary schools in China and has become the most popular Chinese input method on PC platform, the significant difference between Pinyin and the stroke-based method could arguably be attributed to the participants' familiarity with Pinyin. Although Pinyin on mobile phones differs somewhat from the standard PC implementation, the general rules are the same. On the other hand, the stroke-based method was completely new to our participants. We suspect that the 10 minutes of practice that was allowed was not sufficient to allow these individuals to develop an effective mental model of this new solution. While additional practice may allow for improved performance, we believe that design changes could produce more dramatic changes while reducing the time required to become proficient with the stroke-based solution.

PROBLEM IDENTIFICATION

Video recordings of participants interacting with the stroke method during the preliminary comparison were reviewed in detail. This analysis confirmed that the participants pressed the wrong key for more than one third of all strokes they entered. Clearly, the keypad graphics did not provide our participants with the information they needed to map individual strokes to keys, resulting in numerous errors as participants completed the experimental task.

Three types of mapping errors were identified:

- **Correct stroke/incorrect key:** These errors occurred when participants decomposed the character into valid strokes, but selected the wrong key when entering the stroke.
- **Stroke subdivision:** These errors occurred when participants divided a single stroke into more than one piece and tried to enter each piece separately.
- **Stroke combination:** These errors occurred when participants combined multiple strokes and attempted to enter the combination of strokes with a single key press.

These observations indicated that although the character decomposition process that forms the foundation of the stroke method was familiar to participants, the specific symbols printed on the keypad caused confusion and errors. With the current symbols, participants were not able to learn the stroke-to-key mappings. Participants had more difficulty entering strokes associated with the 1, 3, 7, and 9 keys due to the diversity of strokes associated with these keys.

REDESIGNING THE GRAPHICS

The goal for this redesign was to help users better understand the stroke-to-key mappings. All of the strokes entered using any single key share common characteristics, but these common characteristics were not effectively highlighted by the original graphics. Therefore, our first goal was to more effectively convey this information to the user. Abstract symbols were designed for the 1, 3, 7, and 9 keys that highlighted the characteristics that were used to group the strokes. These abstract symbols were designed such that they would not be confused with any specific, real strokes. In addition to displaying the abstract symbols, several example strokes were also presented to highlight the diversity of strokes that are represented by individual keys. Three new designs were generated: abstract, abstract-with-examples, and original-with-examples (Figure 4).

INITIAL EVALUATION OF NEW KEYPAD GRAPHICS

A study was designed to evaluate the effectiveness of the three new designs illustrated in Figure 4 as well as the original design. The goal of this study was to assess how well the various keypad graphics matched users' mental models when they first encountered the keypad. This was assessed by measuring how accurately users could map strokes to specific keys using the various graphics.

Participants

A power analysis based on pilot data suggested that 32 participants would be adequate for a between-group study. Thirty-two native Chinese speakers living in the Maryland area volunteered to participate in this study. Participants were familiar with simplified Chinese and had no difficulties listening to Mandarin. Participants were randomly assigned to use one of the four designs. The gender and age information for each group were listed in Table 2.

Tasks

The experimental task was to enter all of the strokes necessary to input 40 Chinese characters. The characters were carefully selected to ensure that participants would enter all possible strokes at least one time if they completed the task correctly. The 40 characters were presented using audio recordings. To help ensure that participants could determine exactly which character was to be entered, the target character was placed in the context of a two-character phrase. The target character was always the first of the two characters in the phrase. The character set was selected to ensure that all possible strokes would be entered, while keeping the number of strokes required for any given character reasonably low. On average, 4.2 strokes were required to enter each character.

Figure 4. Three new keypad symbol designs: (a) Abstract (b) Abstract-with-examples (c) Original-with-examples

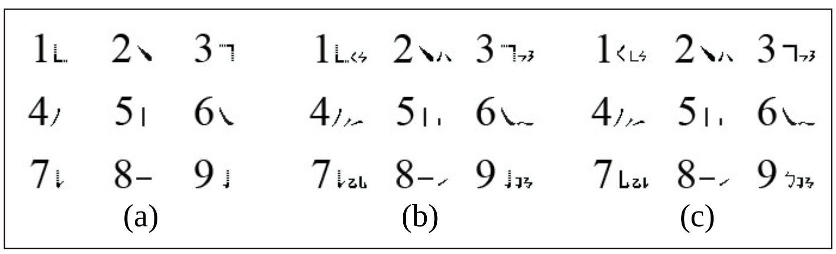
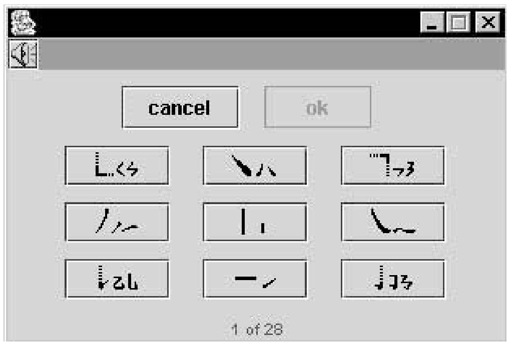


Table 2. The gender and age distribution of the four groups of participants

Design	Gender		Age Range (mean)
	Male	Female	
Original	4	4	23-35 (27)
Abstract	3	5	22-40 (28)
Original-with-examples	3	5	24-45 (30)
Abstract-with-examples	6	2	22-36 (28)

Figure 5. Java application for the blind test (abstract-with-examples design was shown)



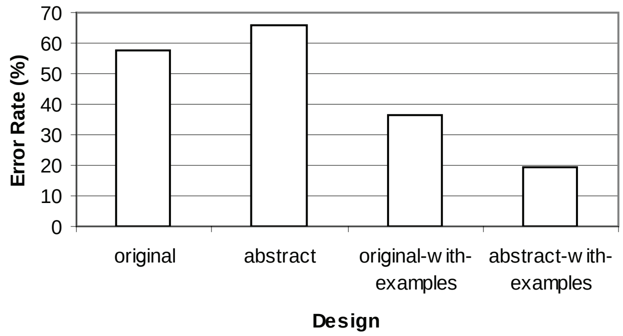
Since the primary goal was to assess the participants' initial mental models regarding the relationship between individual strokes and keys, we designed a Java application (Figure 5) that provides visual feedback regarding the key selected, but no information regarding whether the input would produce the desired character. This ensured that our participants did not learn from their interactions with the application, and that our results reflect the participants' initial mental models regarding the stroke-to-key mappings.

Results

Both character- and stroke-level accuracy were measured. Stroke-level accuracy assessed whether each key corresponded to the desired stroke. Character-level accuracy assessed whether or not the individual could have entered the character correctly using the character prediction technique provided by the iTap™ technique. Using these character prediction capabilities, users typically enter a subset of the strokes required to write the character. After each key is selected, a set of possible strokes is presented, based on the complete sequence of keys that had been entered. Character-level accuracy was assessed by assuming that the participant would scan the list of possible characters after each key is pressed, selecting the desired character as soon as it appeared.

Keypad design had a significant effect on the character-level accuracy [$F(3,28)=21.42$, $p<0.001$]. Post hoc tests revealed no statistical difference between the original design and the abstract design. The original-with-examples design resulted in significantly higher accuracy than the original design [$t(14)=3.35$, $p<0.01$]. Similarly, the abstract-with-examples design resulted in significantly higher accuracy than the abstract design [$t(14)=7.07$, $p<0.001$]. Finally, the abstract-with-examples design resulted in significantly higher accuracy than the original-with-examples design

Figure 6. The character-level error rate (using character prediction)

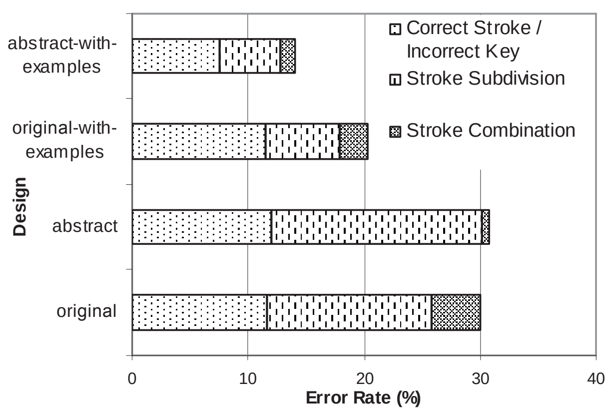


[t(14)=2.61, p<0.05]. The character-level accuracy results confirm that even without feedback regarding the stroke-to-key mappings, the abstract-with-examples solution allowed users to correctly enter almost twice as many characters as compared to the original keypad design (Figure 6).

The stroke-level accuracy was considered more important, since it directly reflects the stroke-to-key mappings that participants develop, given the graphics presented on the keypads. Figure 7 shows overall error rates for the four keypad designs, as well as the frequency of the three different types of errors that occurred. Overall, the pattern was the same as was observed for character-level accuracy: no significant difference was found between the original and abstract designs; adding examples significantly reduced error rates [original-with-examples vs. original: t(14)=-4.39, p<0.002; abstract-with-examples vs. abstract: t(14)=-6.24, p<0.001]; the abstract-with-examples design resulted in fewer errors than the original-with-examples design [t(14)=-2.70, p<0.02].

When looking at the three different types of mapping errors, we found that adding examples reduced the number of errors where a stroke was subdivided and entered using multiple key presses. We also found that both the abstract and the abstract-with-examples designs resulted in significantly fewer errors where participants tried to enter multiple strokes by combining them and using a single key press. Finally, the abstract-with-examples design resulted in fewer errors than any other design that involved situations where participants knew the stroke but selected the wrong key. These results indicated that both the abstract symbols and the example strokes were useful in

Figure 7. The stroke-level error rates using the four keypad designs



reducing mapping errors. Combining the abstract symbols with the example strokes provided additional benefits, reducing the frequency of all three types of errors. A more detailed treatment of this study can be found in Lin and Sears (2007).

HANDS-ON TEST

Our initial evaluation confirmed that combining abstract symbols with examples improved the participants' understanding of stroke-to-key mappings. However, the design of this initial study did not provide feedback regarding the efficacy of the proposed solution when individuals are using a mobile phone to enter text. Therefore, a hands-on longitudinal study was conducted to examine both performance with, and the learnability of, the proposed solution.

Participants

All participants were born and raised in China and had been living in the United States for no more than 4 years. A total of 26 people from the Baltimore/Washington area participated. The first two participants took part in a pilot study that allowed us to determine how many days the actual study would last. The other 24 participants took part in the actual study.

The participants were comfortable understanding Madarin and writing simplified Chinese, but none had experience using iTap™ solutions. Half of the participants had a minimum of a Bachelor's degree while the other half had earned a Master's degree. Participants were randomly assigned into two task groups to use either the original keypad design or the abstract-with-examples design (see table 3).

Tasks

Motorola produced the new abstract-with-examples keypad for this experiment with the same finish as the already-released keypad so that the mobile phone using the new keypad would have the same look and feel as a mature product rather than a prototype. Participants used one of two phones: the standard phone with the original keypad graphics or the new phone with the abstract and example graphics. Our pilot study indicated that 6 days would provide sufficient time for our participants to learn how to use the phones to enter text effectively and for performance to level off.

The primary task for each participant was to use the assigned mobile phone to enter five 17-character sentences. The complete set of 30 sentences was generated based on the headlines of a popular Chinese news Web site, with the topics spread over five categories: international affairs, economy, education, technology, and sports. As a result, each participant entered a total of 510 characters using 361 unique characters during the 6-day study. As in the earlier study, the sentences were presented using audio recordings. After hearing the audio, participants wrote the sentence on paper before entering it using the mobile phone. Participants were instructed to balance the input speed and accuracy as they normally would.

Each day, after entering the five sentences, participants completed a calculation task that consisted of adding 20 sets of three 2-digit numbers. This task was included with the goal of distracting the participants so that they were prevented from rehearsing the stroke-to-key mappings. Since rehearsal is required to retain information in short-term memory, this process ensured that responses during the subsequent

Table 3. The gender and age distribution of the two groups of participants

Keypad	Gender		Age Range (mean)
	Male	Female	
Original	6	6	22-35 (28)
Abstract-with-examples	6	6	23-29 (26)

character entry task were based entirely on the participants' long-term memories. Following the calculation task, participants entered 28 characters using the application from our first study. The characters were selected to include all the strokes at least twice, with the exception of 2 strokes that only exist in one character. The average number of strokes per character was 5.2. The order in which participants entered the characters was randomized each day. This task was included to assess how the participants' mental models of the stroke-to-key mappings evolved as they gained more experience using the phone. At the conclusion of the 6-day study, each participant completed a simple key-pressing task, providing a measure of how fast each individual normally pressed the keys on the mobile phone. This measure of keystroke speed was integrated into our subsequent analyses to address any individual differences with regard to how fast participants normally pressed the buttons.

Results

ANCOVA analysis, using keystroke speed as a covariate, revealed significant effects of keypad design [$F(1, 21) = 5.2$, $p<0.04$] and experiment trial [$F(5, 105) = 10.2$, $p<0.001$] (Figure 8). Both groups entered text more quickly as they gained experience, but participants using the abstract-with-examples design were consistently faster than those using the original design.

During our preliminary evaluation, individuals entered text using Pinyin at a rate of 5.5cpm. Since this is the most frequently used method for entering Chinese text when using a full keyboard (Yuan, 1997), and all participants in the preliminary evaluation were experienced using Pinyin on PCs, we use this as a baseline for comparisons. Our data indicate that with practice, the stroke method did allow for faster data entry than Pinyin. When using the original keypad design, data entry rates surpassed Pinyin on the fourth trial, with a speed of 5.7cpm. Data entry rates for the new abstract-with-examples design exceeded those achieved with Pinyin during the third trial, with an average speed of 6.0cpm. While this appears promising, it is more important to know how much time users must invest before they achieve this level of performance. Figure 9 illustrates the cumulative time spent interacting with the two mobile phones across the six trials. With the original design, individuals invested a total of 95 minutes interacting with the system before they were able to enter text more quickly than the baseline rate for Pinyin. In contrast, only 62 minutes were required to achieve the same goal with the new abstract-with-examples keypad.

An ANOVA analysis with repeated measures showed significant effects of both trial [F(5, 110)

Figure 8. Entry speeds of the two keypad designs

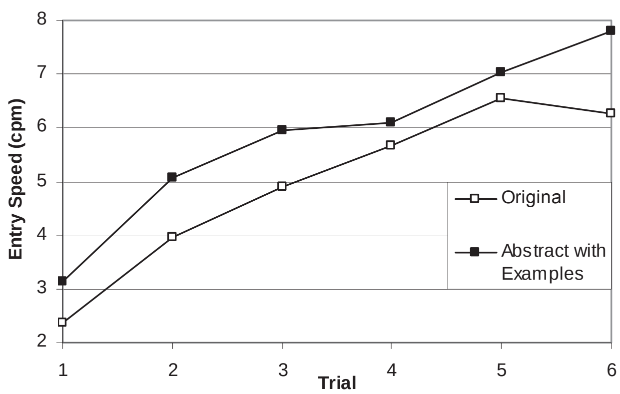


Figure 9. The accumulative time spent for text-entry tasks when using the two designs

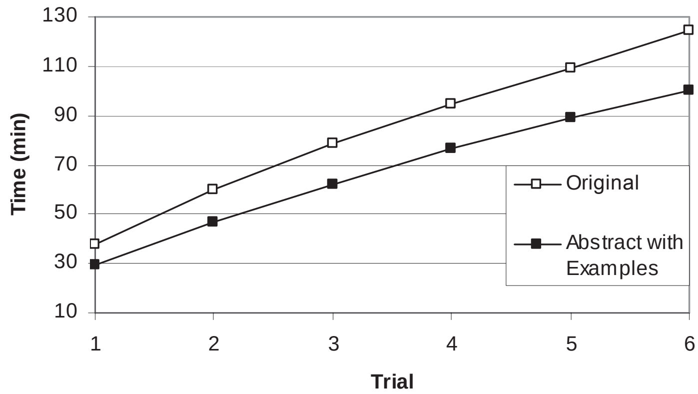
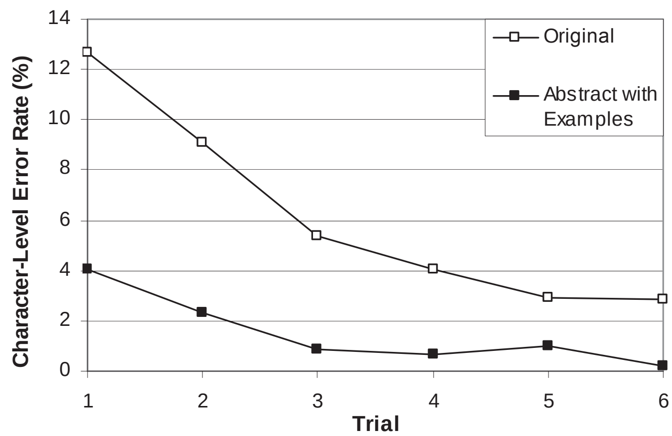


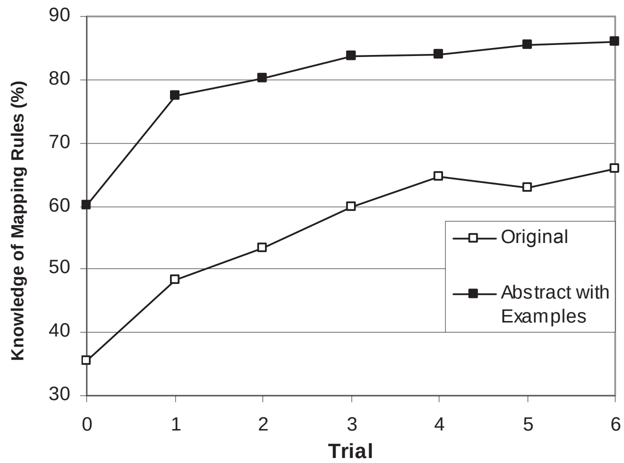
Figure 10. Character-level input error rates when using the two designs



= 16.3, p<0.001] and keypad design [F(1, 22) = 6.7, p<0.02] on the character-level error rates (Figure 10). A significant interaction between trial and keypad design was also detected [F(5, 110) = 3.8, p<0.004], indicating that performance when using the abstract-with-examples design was more stable than performance with the original design. With practice, error rates decreased from almost 13% to approximately 3% for the original design. Interestingly, error rates for the abstract-with-examples keypad started off at just 4% and decreased to 0.2% with practice. Throughout the 6-day period, error rates for the abstract-with-examples design were always lower. In fact, even with 6 days of practice, the error rate for the original design never dropped below the day two error rate for the abstract-with-examples design.

During our preliminary evaluation, the error rate for the stroke-based solution was 7.6%, which was lower than the day-one error rate for the current study. We believe that several factors may have contributed to the increased error rate in the current study. First, unlike the current study, participants in the preliminary study were given 10 minutes to practice using the phone. More importantly, a larger character set was used in the current study as compared to the preliminary study. Even if an error rate of 7.6% were used for comparison, individuals using the new abstract-with-examples design performed much better during their first

Figure 11. Learning curves of the stroke-to-key mappings by the two groups



interactions with the phone (i.e., 4% error rate). When compared to Pinyin, where we observed a 1.5% error rate, users of the new abstract-with-examples design were able to reach a comparable error rate during day three, while error rates when using the original design were still 60% higher than the Pinyin baseline even during day six. These findings strongly supported the use of the new abstract-with-examples design.

While the data entry and error rates were very encouraging, we were also interested in understanding the users' mental models regarding stroke-to-key mappings and how these models evolved with practice. Using the data gathered at the end of each day, we are able to determine how many strokes each individual could map to the correct key. Figure 11 highlights these results, showing how the participants' mental models became more accurate as they interacted with the mobile phones. When using the original design, participants mapped approximately 35% of the strokes to the correct keys before using the phones, almost 50% after using the phones just once, and stabilized around 65% after day four. In contrast, when using the new design, participants mapped 60% of the strokes correctly before using the phone, nearly 80% after using the phones once, and stabilized around 85% after day three.

Significant effects of trial [$F(5, 110) = 18.6$, $p<0.001$] and keypad design [$F(1, 22) = 30.0$, $p<0.001$] on the understanding of stroke-to-key mappings were observed. The interaction between trial and keypad design was also significant [$F(5, 110) = 2.6$, $p<0.03$], consistent with the character-level error analysis (Figure 10). Regressions between the knowledge of mapping rules and character-level error rates confirmed strong linear correlations. The regression coefficients for the original design and the abstract-with-examples design were 0.97 ($p<0.001$) and 0.94 ($p<0.002$), respectively. Therefore, reductions in character-level error rates can be associated with improved knowledge of the stroke-to-key mappings. A more detailed description of this study can be found in Lin and Sears (2005a).

CONFIRMATION TEST

The longitudinal study described confirmed that the new abstract-with-examples design allowed for more rapid learning, faster data entry, and lower error rates. While the participants were all from China, they were currently living in the United States. Therefore, an additional study was conducted in China to ensure that the results were not inappropriately biased by the participant's current location or the culture in which they were currently living. The instructions and materials from the previous study were used to conduct this

Table 4. The gender and age distribution of the two groups of participants

Keypad	Gender		Age Range (mean)
	Male	Female	
Original	6	6	22-36 (27)
Abstract-with-examples	6	6	23-36 (27)

follow-up study to ensure that the results could be compared.

Participants

Twenty-four individuals in Beijing were recruited to participate in the study (table 4). As with the US-based study, half of the participants had a Bachelor's degree and the other half had earned a Master's degree. Participants were randomly assigned to two groups, one using the original keypad and the other using the new abstract-with-examples design.

Results

The patterns observed in the US-based study were also apparent in the results from the follow-up study in China. Figure 12 shows the data-entry speeds achieved over the 6-day study.

The character-level error rates also resulted in a similar pattern (Figure 13). The results for the abstract-with-examples design were almost identical to those from the US-based study. Error rates for the original design were somewhat lower than in the US-based study, but the pattern was similar. As in the US-based study, error rates for the original keypad were always higher than those for the new abstract-with-examples design.

SIMPLIFICATION TEST

The level of consistency between the US- and China-based studies was encouraging, confirming the robustness of our results. While these results reaffirmed the belief that the abstract-with-examples design should replace the original design, one practical challenge still had to be addressed: mobile phones in the Asian market tend to be small and continue to get smaller. As a result, the keys on these phones are quite small, and may make it difficult to fit both the abstract symbol and two examples on each key. To ensure that the end result addressed both usability and business goals, we studied the possibility of simplifying the abstract-with-examples design. More specifically, we focused on the possibility of using just a single example stroke.

Figure 12. Data entry speeds using the two keypad designs when repeated in China

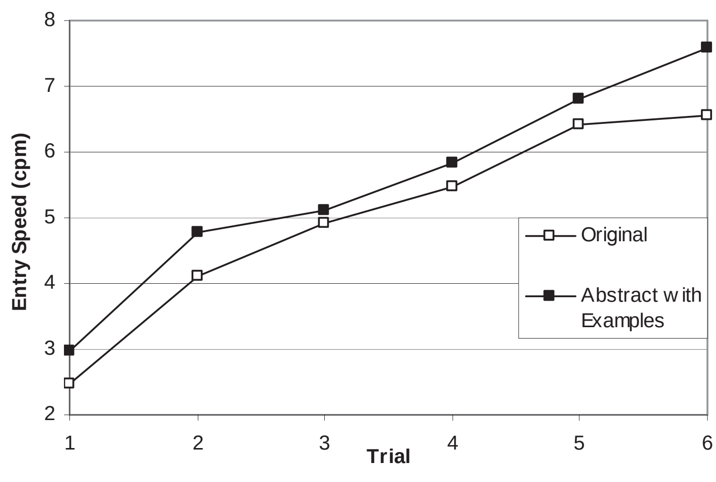
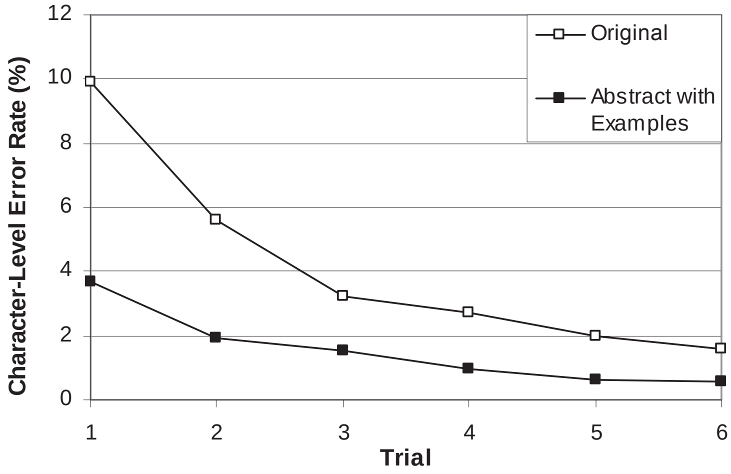


Figure 13. Character-level input error rates when repeated in China



As described earlier, the primary goal for the abstract-with-examples keypad design was to enhance the users' understanding of the stroke-to-key mappings for four specific keys. Our simplification efforts also focused on these four keys (i.e., 1, 3, 7, and 9). Each key originally had two example strokes. In the current study, two alternative designs were considered for each of these keys. The alternatives paired the abstract symbol with one of the example strokes used in the initial abstract-with-examples design with the goal of determining which example stroke produced better results.

Table 5. The gender and age distribution of the eight groups of participants

Group	Gender		Age Range (mean)
	Male	Female	
1 (key 1 design #1)	6	14	22-35 (27)
2 (key 1 design #2)	15	5	24-37 (29)
3 (key 3 design #1)	11	9	24-34 (28)
4 (key 3 design #2)	9	11	18-33 (27)
5 (key 7 design #1)	8	12	21-34 (27)
6 (key 7 design #2)	7	13	22-34 (28)
7 (key 9 design #1)	7	13	22-37 (27)
8 (key 9 design #2)	10	10	25-41 (31)

Participants

One hundred and sixty participants were recruited in Beijing (Table 5). Participants were randomly assigned to one of the eight groups, with each group evaluating one design for a single key.

Tasks

A Web-based study was used, with each participant being assigned a unique user id and password to limit access to the experimental materials. Each user id and password worked one time, ensuring that participants did not complete the study more than once. Thirty six characters were presented to the participants one at a time. A single stroke was highlighted in each character. The set of 36 characters ensured that every possible stroke was presented one time. For each character, the participant was also presented a single key design that was under consideration. The specific key design presented was determined based on the group the participant had been assigned to. The participant had to judge whether the design should be used to enter the highlighted stroke.

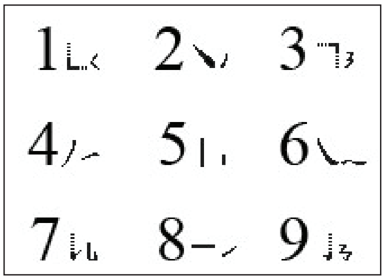
Results

Since each key was presented in isolation, participants had to make decisions without the insights

Table 6. Error rates of the four pairs of one-example design alternatives

Design	Key 1		Key 3		Key 7		Key 9	
	#1	#2	#1	#2	#1	#2	#1	#2
Error rate (%)	68	76	74	68	82	78	82	76

Figure 14. Abstract-with-one-example design



provided by the eight additional keys that would be available when interacting with a full keypad. The current task is further complicated by the fact that during character entry tasks, individuals judge which one of the nine keys is most appropriate, but in the current task they must determine if a specific stroke matches a specific key. Given the more complex nature of the current task, we expected stroke-level error rates to be substantially higher than would be experienced during the use of a real phone keypad. The mean stroke-level error rates are reported in table 6. Using paired t-tests, we did not identify any significant differences between the pairs of alternative designs for any of the four keys in question. While the differences were not statistically significant, we chose to use the design with the lower error rate for use in the abstract-with-one-example design (Figure 14).

FINAL EFFECTIVENESS TEST

As discussed earlier, new mobile phones developed by Motorola for the Chinese market continue to get smaller, severely limiting the amount of space available on individual keys. To balance usability goals and the reality of having very limited space available on each key for graphics, our last study focused on simplifying the abstract-with-examples solution to reduce the space required for the keypad graphics. The final evaluation focused on assessing the impact of the abstract-with-one-example design on the users' ability to determine correct stroke-to-key mappings. For this study, three designs were evaluated: the original keypad, the new abstract-with-examples design, and the abstract-with-one-example design.

Participants

Sixty participants were recruited in Beijing and randomly assigned to one of three groups (table 7). Each group interacted with a single keypad design.

Tasks

The 36 characters used in the previous test were used once again in this final evaluation that used the similar Web-based approach as the last study. When a character was presented the full keypad was displayed, and the participant had to indicate which key would be used to enter the highlighted stroke.

Table 7. Gender and age distribution of the three groups

Keypad	Gender		Age Range (mean)
	Male	Female	
Original	8	12	21-40 (30)
Abstract-with-examples	7	13	23-38 (29)
Abstract-with-one-example	8	12	26-34 (30)

Figure 15. The overall error rates for the three keypad designs

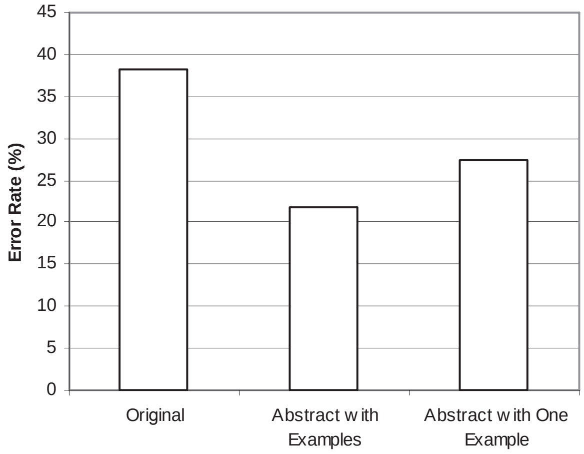


Figure 16. Error rates for the four critical keys

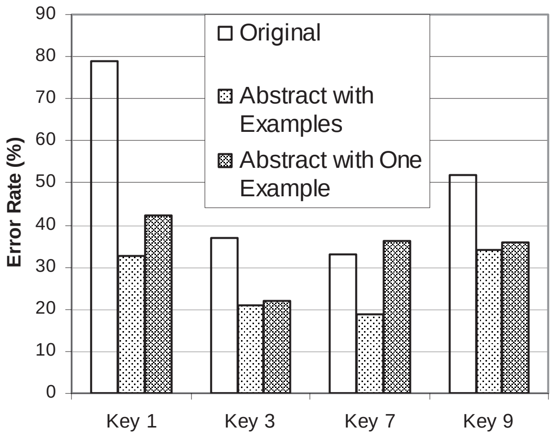


Figure 17. Keypads of the Motorola A732 and E2 phones (showing the abstract-with-one-example design). Both images © 2007, Motorola, Inc. Reproduced with Permission from Motorola, Inc.

Figure 18. Keypads of the Motorola K1 and V3xx phones (showing the abstract-with-one-example design). Both images © 2007, Motorola, Inc. Reproduced with Permission from Motorola, Inc.

Results

One outlier was identified in each group using a QQ plot. Outliers were excluded from the subsequent analyses, which revealed a significant effect of group on stroke-to-key mapping error rates (Figure 15) [$F(2, 54) = 13.6$, $p<0.001$]. Group also had a significant effect on the error rates for the 1, 3, and 9 keys [$F(2, 54) = 32.6$, $p<0.001$; $F(2, 54) = 4.2$, $p<0.021$; $F(2, 54) = 3.6$, $p<0.034$].

Figure 16 shows the error rates of the four critical keys that were changed when the simplified design was developed. For 1, 3, and 9 keys, the error rate for the original design was significantly higher than for the abstract-with-examples design or the abstract-with-one-example design. At the same time, no significant differences were detected between the abstract-with-examples design and the abstract-with-one-example design for these three keys. These findings indicate that the abstract-with-one-example design provides benefits that are comparable to the abstract-with-examples design even after one example stroke has been removed. In contrast, removing one example from the abstract-with-examples solution for the 7 key resulted in the error rate increasing back to the level observed for the original design. We believe that this is due, in large part, to the fact that this key represents the largest number of strokes, and that the diversity of these strokes is difficult to capture with just a single example. Figures 17 and 18 provide several examples of Motorola phones that have adopted the new keypad graphics and are now available in various markets.

CONCLUSION

In this case study, we presented the step-by-step development of a new keypad design for use with Motorola's iTap™ stroke input technique. Most current research on Chinese character input focuses on key sequence optimization, including various methods of character prediction. The current study is unique in that it focused on the process of entering individual strokes more efficiently by helping users more effectively map strokes to specific inputs. The success of this project is due to the close collaboration between UMBC and Motorola personnel in both the US and China. Motorola helped motivate the problem and UMBC completed many of the initial studies with guidance and input from Motorola personnel. Motorola then lead the effort to replicate the UMBC studies in China, with UMBC providing guidance and developing new studies to address new issues such as simplifying the keypad designs. UMBC then worked with Motorola to develop the necessary materials that would allow for an informed decision regarding the adoption of this new design. Motorola personnel completed the final, internal steps to address certain questions that were raised during the process of formally adopting the new keypad graphics.

The resulting solutions are cost-effective, since the only change is to the graphics that are printed on the keypads, but they are also effective in that they allow for rapid learning, more efficient text entry, and reduced errors. More importantly, the new graphics make the stroke-based method a competitive alternative to the popular pronunciation-based input method (i.e., Pinyin), especially for individuals who do not speak Mandarin or speak Mandarin in a nonstandard way. Our studies confirmed that approximately 1 hour of casual practice with the new keypad is sufficient to allow for text-entry and error rates that are comparable or superior to those that have been observed with Pinyin. Mobile phones that use these new designs began shipping in the Chinese market in 2006.

ACKNOWLEDGMENT

We want to thank Mr. Jun Chen from Motorola China. Mr. Chen helped conduct the usability testing that examined the effectiveness of simplified design. This research was sponsored by Motorola.

REFERENCES

Archer, N. P., Chan, M. W. L., Huang, S. J., & Liu, R. T. (1988). A Chinese-English microcomputer system. *Communications of the ACM*, *31*(8), 977-982.

Census and Statistics Department, Hong Kong Special Administration Region, PRC (2001). *Population of five-year-old and above, categorized by primary speaking language in year 1991, 1996, and 2001.* Retrieved April 15, 2004, from http://www.info.gov.hk/censtatd/chinese/hkstat/fas/01c/cd0062001c_text.htm

Cockburn, A., & Siresena, A. (2003). Evaluating mobile text entry with the Fastap keypad. In *Proceedings of British Computer Society Conference on Human Computer Interaction* (pp. 77-80). London: Springer-Verlag.

James, C. L., & Reischel, K. M. (2001). Text input for mobile devices: Comparing model prediction to actual performance. In *Proceedings of CHI 2001* (pp. 365-371). New York: ACM Press.

Lin, M., & Sears, A. (2005a). Chinese character entry on mobile phones: A longitudinal investigation. *Interacting with Computers, 17,* 121-146.

Lin, M., & Sears, A. (2005b). Graphics matter: A case study of mobile phone keypad design for Chinese input. In *Proceedings of CHI 2005* (pp. 1593-1596). New York: ACM Press.

Lin, M., & Sears, A. (2007). Constructing Chinese characters: Keypad design for mobile phones. *Behavior and Information Technology.*

MacKenzie, I. S., Kober, H., Smith, D., Johns, T., & Skepner, E. (2001). LetterWise: Prefix-based disambiguation for mobile text input. In *Proceedings of UIST 2001* (pp. 111-120). New York: ACM Press.

Matić, N. P., Platt, J. C., & Wang, T. (2002). QuickStroke: An incremental on-line Chinese handwriting recognition system. In *Proceedings of 16th International Conference on Pattern Recognition,* vol. 3 (pp. 435-439). New York: IEEE Computer Society Press.

Pavlovych, A., & Stuerzlinger, W. (2004). Model for non-expert text entry speed on 12-button phone keypads. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 351-358). New York: ACM Press.

Sacher, H., Tng, T.-H., & Loudon, G. (2001). Beyond translation: Approaches to interactive products for Chinese consumers. *International Journal of Human-Computer Interaction, 13*(1), 41-51.

Silfverberg, M., MacKenzie, I. S., & Korhonen, P. (2000). Predicting text entry speed on mobile phones. In *Proceedings of CHI 2000* (pp. 9-16). New York: ACM Press.

Wigdor, D., & Balakrishnan, R. (2003). TiltText: Using tilt for text input to mobile phones. In *Proceedings of the 16th Annual ACM Symposium on User Interface Software and Technology* (pp. 81-90). New York: ACM Press.

Yuan, C. (1997). *Chinese language processing.* Shanghai, China: Shanghai Education Publishing Company.

KEY TERMS

Chinese Character: A Chinese character is the minimum functional unit of Chinese language.

iTap™ Software: iTap™ software was developed by Lexicus and Motorola, and enables predictive text entry in mobile phones. iTap™ software supports both Pinyin and stroke-based entry.

Mandarin: Mandarin is a northern Chinese dialect that is the basis for the official pronunciation of each Chinese character for Pinyin.

Mental Model: A mental model is the users' internal representation of how a system works.

Pinyin: Pinyin is the official Romanization system for Mandarin. It uses the 26 letters of the Roman alphabet to define the pronunciation of Chinese characters so that they can be entered using the standard western keyboard.

Stroke: Stroke is the minimum writing unit of Chinese language. Each Chinese character is constructed by writing one or more strokes in a specific order while following specific spatial relations.

Text Entry: Text entry refers to the process of creating messages composed of characters, numbers, and symbols using mobile devices. Text entry can be performed using small physical keys, virtual keyboards presented on touch sensitive screens, gesture or handwriting recognition, speech recognition, and various other technologies.

Chapter XXVII
Voice-Enabled User Interfaces for Mobile Devices

Louise E. Moser
University of California, Santa Barbara, USA

P. M. Melliar-Smith
University of California, Santa Barbara, USA

ABSTRACT

The use of a voice interface, along with textual, graphical, video, tactile, and audio interfaces, can improve the experience of the user of a mobile device. Many applications can benefit from voice input and output on a mobile device, including applications that provide travel directions, weather information, restaurant and hotel reservations, appointments and reminders, voice mail, and e-mail. We have developed a prototype system for a mobile device that supports client-side, voice-enabled applications. In fact, the prototype supports multimodal interactions but, here, we focus on voice interaction. The prototype includes six voice-enabled applications and a program manager that manages the applications. In this chapter we describe the prototype, including design issues that we faced, and evaluation methods that we employed in developing a voice-enabled user interface for a mobile device.

INTRODUCTION

Mobile devices, such as cell phones and personal digital assistants (PDAs), are inherently small, and lack an intuitive and natural user interface. The small keyboards and displays of mobile devices make it difficult for the user to use even the simplest of applications. Pen input is available on PDAs, but is difficult to use on handheld devices.

Voice input and output for mobile devices with small screens and keyboards, and for hands- and eyes-free operation, can make the user's interaction with a mobile device more user friendly. Voice input and output can also facilitate the use of Web Services (Booth, Hass, McCabe, Newcomer, Champion, Ferris, & Orchard, 2004) from a mobile device, making it possible to access the Web anytime and anywhere, whether at work, at home, or on the move. Global positioning system (GPS) technology (U.S. Census Bureau, 2006) can provide location information automatically for location-aware services.

Many everyday applications can benefit from voice-enabled user interfaces for a mobile device. Voice input and voice output for a mobile device are particularly useful for:

- Booking theater and sports tickets, making restaurant and hotel reservations, and carrying out banking and other financial transactions
- Accessing airline arrival and departure information, weather and traffic conditions, maps and directions for theaters, restaurants, gas stations, banks, and hotels, and the latest news and sports scores
- Maintaining personal calendars; contact lists with names, addresses, and telephone numbers; to-do lists; and shopping lists
- Communicating with other people via voice mail, e-mail, short message service (SMS), and multimedia message service (MMS).

It is important to provide several modes of interaction, so that the user can use the most appropriate mode, depending on the application and the situation. The prototype system that we have developed supports client-side, voice-enabled applications on a mobile device. Even though the applications support multimodal input, allowing keyboard and pen input, we focus, in this chapter, on voice input and on multimodal output in the form of voice, text, and graphics. The prototype includes a program manager that manages the application programs, and six voice-enabled applications, namely, contacts, location, weather, shopping, stocks, and appointments and reminders.

BACKGROUND

A multimodal interface for a mobile device integrates textual, graphical, video, tactile, speech, and/or other audio interfaces in the mobile device (Hjelm, 2000; Oviatt & Cohen, 2000). With multiple ways for a user to interact with the applications, interactions with the device become more natural and the user experience is improved. Voice is becoming an increasingly important mode of interaction, because it allows eyes- and hands-free operation. It is essential for simplifying and expanding the use of handheld mobile devices. Voice has the ability to enable mobile communication, mobile collaboration, and mobile commerce (Sarker & Wells, 2003), and is becoming an important means of managing mobile devices (Grasso, Ebert, & Finin, 1998; Kondratova, 2005).

The increasing popularity of, and technological advancements in, mobile phones and PDAs, primarily mobile phones, is leading to the development of applications to fulfill expanding user needs. The short message service (SMS) is available on most mobile phones today, and some mobile phones provide support for the multimedia messaging service (MMS) to exchange photos and videos (Le Bodic, 2002). The mobile phone manufacturers are no longer focused on making a mobile phone but, rather, on producing a mobile device that combines phone capabilities with the power of a handheld PC. They recognize that the numeric keypad and the small screen, common to mobile phones of the past, do not carry over well to handheld PCs (Holtzblatt, 2005).

With the emergence of Web Services technology (Booth et al., 2004), the Web now provides services, rather than only data as it did in the past. Of the various Web Services available to mobile users today, the map application seems to be the most popular, with online map services available from Google (2006) and Yahoo! (2006b). Much progress has been made in creating the multimodal Web, which allows not only keyboard and mouse navigation but also voice input and output (Frost, 2005).

GPS technology (U.S. Census Bureau, 2006) already exists on many mobile devices, and can be used to provide location-aware services (Rao & Minakakis, 2003), without requiring the user to input geographical coordinates, again contributing to user friendliness.

Speech recognition technology (Rabiner & Juang, 1993) has been developed over many years, and is now very good. Other researchers (Kondratova, 2004; Srinivasan & Brown, 2002) have discussed the usability and effectiveness of a combination of speech and mobility. Currently, handheld voice-enabled applications use short commands that are translated into functional or navigational operations. As observed in Deng and Huang (2004), speech recognition technology must be robust and accurate, and close to human ability, to make its widespread use a reality. Noisy environments present a particular challenge for the use of speech recognition technology on mobile devices and, therefore, multimodal interactions are essential. For example, the MiPad system (Deng, Wang, Acero, Hon, Droppo, Boulis, et al., 2002; Huang, Acero, Chelba, Deng, Droppo, Duchene, Goodman, et al., 2001) uses a strategy where the user first taps a "tap & talk" button on the device and then talks to the device.

Distributed speech recognition (Deng, et al., 2002), in which the speech recognition happens at a remote server exploits the power of the server to achieve fast and accurate speech recognition. However, studies (Zhang, He, Chow, Yang, & Su, 2000) have shown that low-bandwidth connections to the server result in significant degradation of speech recognition quality. In contrast, *local speech recognition* (Deligne, Dharanipragada, Gopinath, Maison, Olsen, & Printz, 2002; Varga, Aalburg, Andrassy, Astrov, Bauer, Beaugeant, Geissler, & Hoge, 2002) utilizes speech recognition technology on the mobile device, and eliminates the need for high-speed communication. Local speech recognition limits the kinds of client handsets that are powerful enough to perform complicated speech processing and, thus, that can be used; however, the computing power of mobile handsets is increasing.

THE PROTOTYPE

The prototype that we have developed allows mobile applications to interact with the user without the need for manual interaction on the part of the human. Speech recognition and speech synthesis software are located on the mobile device, and make the interaction with the human more user friendly. The prototype that we have developed processes natural language sentences and provides useful services while interacting with the user in an intuitive and natural manner. A user need not form a request in a particular rigid format in order for the applications to understand what the user means.

For our prototype, we have developed six application programs and a Program Manager. These applications are Contacts, Location, Weather, Shopping, Stocks, and Appointments and Reminders applications. The Program Manager evaluates sentence fragments from the user's request, determines which application should process the request, and forwards the request to the appropriate application.

The prototype is designed to interact with a human, using voice as the primary means of input (keyboard, stylus, and mouse are also available but are less convenient to use) and with voice, text, and graphics as the means of output. The speech recognizer handles the user's voice input, and both the speech synthesizer and the display are used for output. Characteristics of certain applications render a pure voice solution infeasible. For example, it is impossible to convey the detailed contents of a map through voice output. However, voice output is ideal when it is inconvenient or impossible for the user to maintain visual contact with the display of the mobile device, and it is possible to convey information to the user in that mode. Voice output is also appropriate when the device requests confirmation from the user.

Thus, an appropriate choice of speech recognition and speech synthesis technology is vital to the success of our prototype. Our choices were constrained by:

- The processing and memory capabilities of typical mobile devices
- The need for adaptability to different users and to noisy environments

The use of speech recognition and speech synthesis technology on a mobile device is different from its use in call centers, because a mobile device is associated with a single user and can learn to understand that particular user.

The Underlying Speech Technology

The prototype uses SRI's DynaSpeak speech recognition software (SRI, 2006) and AT&T's Natural Voices speech synthesis software (AT&T, 2006). It currently runs on a handheld computer, the OQO device (OQO, 2006). We chose this device, rather than a cell phone, because it provides a better software development environment than a cell phone.

Speech Recognition

The DynaSpeak speech recognition engine (SRI, 2006) is a small-footprint, high-accuracy, speaker-independent speech recognition engine. It is based on a statistical language model that is suitable for natural language dialog applications. It includes speaker adaptation to increase recognition accuracy for individuals with different accents or tone pitches. It can be configured so that it performs speech recognition specific to a particular individual. DynaSpeak is ideal for handheld mobile devices, because of its small footprint (less than 2 MB of memory) and its low computing requirements (66 MHz Intel x86 or 200 MHz Strong Arm processor).

DynaSpeak supports multiple languages, adapts to different accents, and does not require training prior to use. It incorporates a Hidden Markov Model (HMM) (Rabiner & Juang, 1993). In an HMM, a spoken expression is detected as a sequence of phonemes with a probability associated with each phoneme. A probability is also associated with each pair of phonemes, that is, the probability that the first phoneme of the pair is followed by the second phoneme in natural speech. As a sequence of phonemes is processed, the probability of each successive phoneme is combined with the transition probabilities provided by the HMM. If the probability of a path through the HMM is substantially greater than that of any other path, the speech recognizer recognizes the spoken expression with a high level of confidence. When the response is below an acceptable confidence threshold, the software seeks confirmation from the user or asks the user questions.

The HMM is augmented with grammars for the particular applications that are required for understanding natural language sentences (Knight, Gorrell, Rayner, Milward, Koeling, & Lewin, 2001). When the user says a new word, the word can be added to the vocabulary dynamically. The HMM is also extended by adapting the vocabulary of the speech recognizer to the current and recent past context of interactions of the user with the applications.

Accuracy of the speech recognition system can be increased by training it for the voice of the particular user. There are two kinds of training, explicit and implicit. *Explicit training* requires the user to read a lengthy script to the device, a process that is likely to be unpopular with users. *Implicit training* allows the device to learn to understand better its particular user during normal use. Implicit training can be provided in two modes, confirmation mode and standard mode.

In *confirmation mode*, the system responds to a user's sentence, and the user confirms or corrects the response. If the user corrects the sentence, the learning algorithm tries to match a rejected, lower probability, interpretation of the original sentence with the user's corrected intent. If a match is found, the learning algorithm adjusts the HMM transition probabilities to increase the probability of selecting the user's intent. Initially, a new user of the system will probably prefer confirmation mode.

In *standard mode*, the system does not confirm sentences for which there is one interpretation that has a much higher probability than any other interpretation. If no interpretation has a high probability, or if several interpretations have similar probabilities, the speech recognition system responds as in confirmation mode. More experienced users of the system are likely to use standard mode.

The success of implicit training strategies depends quite heavily on starting with a speech recognizer that is well matched to the individual speaker. It is possible, from relatively few sentences, to classify a speaker and then to download, to the mobile device, an appropriate initial recognizer for subsequent implicit training.

DynaSpeak can be used with either a *finite-state grammar* or a *free-form grammar*. We used the finite-state grammar because it offers greater control over parsed sentences. The tendency for DynaSpeak to accept or reject spoken sentences is heavily influenced by the complexity of the grammar. The *complexity of the grammar* is quantified by the number of paths by which an accepting state can be reached. The greater the complexity of the grammar, the higher is its tendency to accept an invalid spoken request. Conversely, the lower the complexity of the grammar, the higher is its tendency to reject a valid spoken request. To minimize the complexity of the grammar and to improve speech recognition accuracy, each application has its own relatively simple grammar. The program manager determines which applications are involved in a sentence and then reparses the sentence using the appropriate grammars.

Speech Synthesis

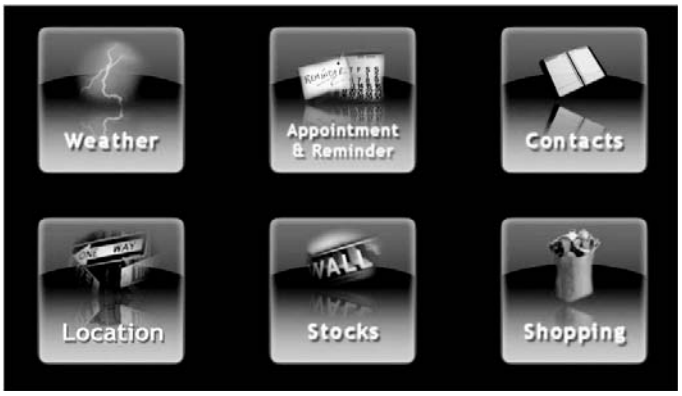
Natural Voices (AT&T. 2006) is a speech synthesis engine that provides a simple and efficient way of producing natural (rather than electronic) sounding device-to-human voice interactions. It can accurately and naturally pronounce words and speak in sentences that are clear and easy to understand, without the feeling that it is a computer that is speaking.

Natural Voices supports many languages, male and female voices, and the VoiceXML, SAPI, and JSAPI interface standards. Using Natural Voices, we created text-to-speech software for our prototype that runs in the background and accepts messages in VoiceXML format. Each message contains the name of the voice engine (i.e., “Mike” for a male voice and “Crystal” for a female voice) and the corresponding text to speak.

Managed Applications

For the prototype we developed six multimodal applications (contacts, location, weather, shopping, stocks, appointments, and reminders) that use speech as the main form of input. The stocks, maps, and weather applications exploit existing Web Services on the Internet. Communication with those Web Services uses a local WiFi 802.11 wireless network. The program manager controls the operation of the applications. The graphical user interface for the program manager with the six applications is shown in Figure 1. We now present an explanation of the functionality of each application and its role in the overall system.

Figure 1. The GUI of the program manager, showing six applications



Contacts

The contacts application stores personal information regarding friends and acquaintances in a database, including their addresses and phone numbers. The contacts application is a mobile extension of a physical contact list or address book that is controlled by voice input. It retrieves data from Microsoft Office Outlook® to populate the database when in docking mode. After using the mobile device and possibly entering new contact information, the user can synchronize information on the mobile device with that on a desktop or server computer. The contacts application is configured to interact with other applications that require information about names, addresses, phone numbers, and so forth. The contacts grammar is the least complex of the application grammars that we developed. The contacts vocabulary grows linearly as contacts are added to the user's contact list.

Location

The Location application allows the user to search for restaurants, movie theaters, banks, and so forth, in a given area, using the Yahoo! LocalSearch Web Service (2006b). For example, if the user says to the mobile device "Search for a Mexican restaurant in 95131," the location application on the mobile device sends a Web Service request to Yahoo! LocalSearch, gets back the results, and presents up to 10 results to the user in list form. The user can then view additional information about a single location by indicating the location's number in the presented list. For example, the user can choose to view additional information about "Chacho's Mexican Restaurant" by speaking, "Get more information about number one." On processing this request, the location application presents the user with detailed information about the restaurant including its phone number, address, and a detailed street map showing its location. Figure 2 shows a screen shot of the graphical user interface for the location application.

The location application is loosely coupled with the contacts application to provide responses related to individuals listed in the user's contact list. For example, the request, "Search for a movie theater around Susan's house" uses the contacts grammar to determine the location of Susan's house and replaces the phrase "Susan's house" with the specific address so that the actual search request looks something like this: "Search for a movie theater around 232 Kings Way, Goleta, CA, 93117." The location application then searches for a movie theater in the vicinity of that address.

The location application is also loosely coupled with a GPS module that is contacted when the user has a question related to the user's current location. For example, if the user says "Look for a pizza place around here.", the word "here" is recognized by the application and replaced with the GPS coordinates of the user's current location.

Figure 2. An example graphical user interface for the location application

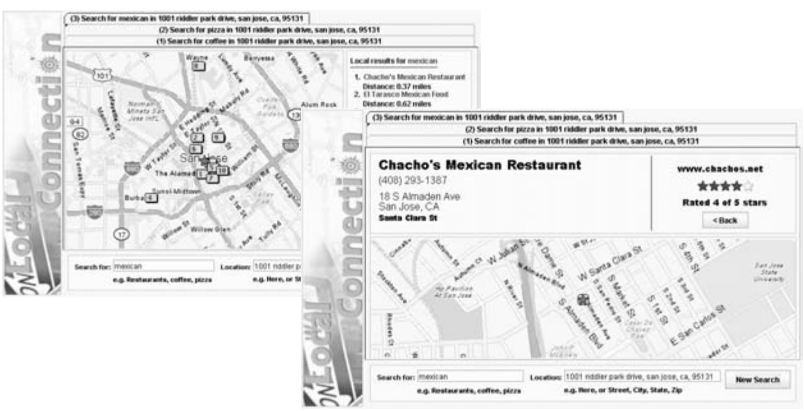
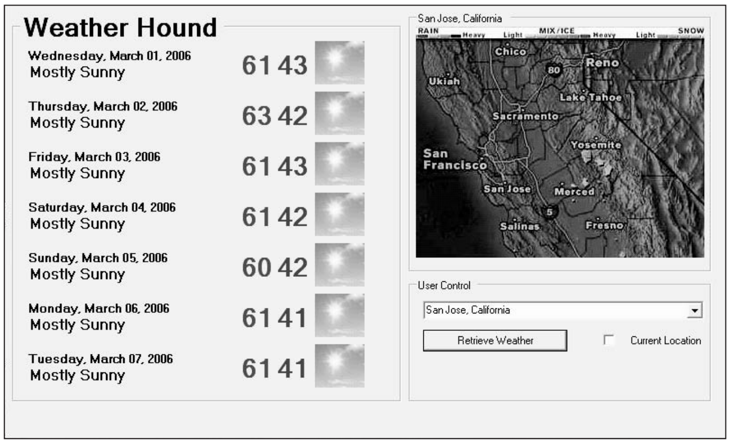


Figure 3. An example graphical user interface for the weather application



The location application then sends a Web Service request to Yahoo! LocalSearch, which returns a map of the user's current location, indicating where the user is, along with the 10 nearest pizza places. The Yahoo! LocalSearch Web Service is ideal to use with GPS because of its ability to locate positions on the map on the basis of longitude and latitude. With GPS, the user is no longer limited to requests involving a particular city or zip code. The user now has the ability to create requests that are truly location-aware.

Compared to the grammars of the other applications, the location grammar is one of the most complex. For information like maps and lists, it is desirable to use a graphical or textual display, as well as speech output, in a multimodal user interface. Thus, the most appropriate kind of output can be chosen, depending on the kind of information, the capabilities of the mobile device, and the context in which the user finds himself or herself.

Weather

The weather application supplies weather forecasts obtained from the Web Service provided by the National Weather Service (NOAA, 2006). It allows the user to query for weekly, daily, and 3-day weather information in major U.S. cities using voice input. It allows the user either to select a city or to use the user's current location, as the location for which the weather forecast is to be retrieved from the National Weather Service. The weather application knows the geographical coordinates of dozens of cities in the continental United States. It references those coordinates when the user requests a weather forecast from the National Weather Service for one of those cities.

A user can say "Tell me the weather forecast in San Jose," which then uses "today" as the starting time of the forecast, and produces the graphical user interface for the weather application shown in Figure 3.

Because the weather application operates on a mobile device, it is necessary to be able to determine the user's location dynamically. If the user asks "What's the weather like here two days from now?", the weather application consults the GPS module to obtain the geographical coordinates of the user, contacts the Web Service, and responds with the high and low predicted temperatures and an indication that there is a change to cloudy in Santa Barbara. Thus, the user does not need to provide his/her current location or to obtain the weather forecast for that location.

Our prototype takes into account the many ways in which a person can convey, semantically, equivalent requests in English. For example, a user can ask for the weather in many ways including "What is the weather in Boston like?" or "Tell me what the forecast is like in Boston." These two requests are semantically equivalent because they both contain the same essential parameter, namely the Boston location.

Shopping

The shopping application provides the user with a service capable of reducing the time that the user spends on grocery shopping and the associated stress. The shopping application maintains shopping lists, recipes, and floor plans of supermarkets. The multimodal interface includes speech, text, and graphics, which makes the shopping application easy to use. Figure 4 shows a screen shot of the graphical user interface for the shopping application.

The shopping application allows a user to update his/her shopping list and to forward it to another user. When a user issues a command, like “Remind John to go grocery shopping,” the contacts application is used to find John’s phone number or e-mail address in the user’s contact list. A dialog box then appears asking the user if he/she wants to send, to John, not only a reminder to go shopping but also the shopping list. If so, the shopping list, consisting of the product ids and the quantities of the items needed, is formatted in XML, and appended to the message containing the reminder. The message is then sent to John’s shopping application.

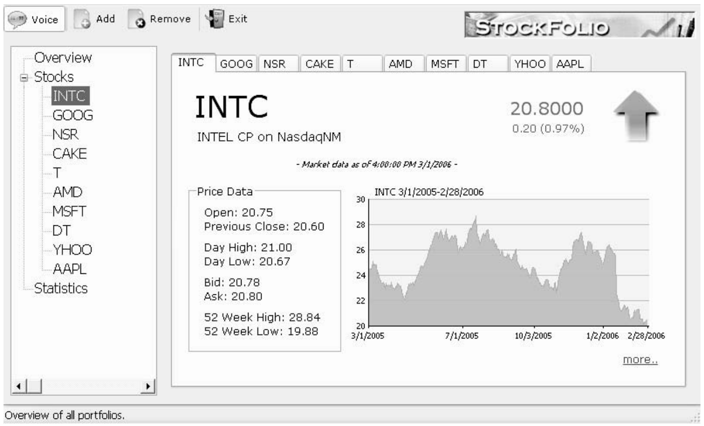
The shopping application also displays graphically the floor plan of the supermarket and the location of items in the store, as shown in Figure 4. This feature provides assistance to the user without the need for the user to contact an employee of the supermarket. The shopping application also allows the user to retrieve recipes while shopping, possibly on impulse, for an item that is on sale. A newly chosen recipe is cross-referenced with the current shopping list, so that needed items can be added automatically. The shopping application has the largest grammar of the applications that we developed, with a vocabulary that depends on the items that the user has purchased recently.

Stocks

The stocks application allows the user to manage his/her stock portfolio using voice input and output. The objective of the stocks application is to monitor stock fluctuations, rather than to trade stocks. The stocks application exploits the Yahoo! Finance Web service (2006a) to store and update stock information in a database. It stores the most recent stock information in the database so that it can reply to the user’s requests when connectivity to the Yahoo! Finance Web Service is limited. Although such stored data can be somewhat stale, it allows the user to obtain information whenever the user requests it. The vocabulary of the stocks application grows to match the user’s portfolio each time the user adds a new stock.

Figure 4. An example graphical user interface for the shopping application

Figure 5. An example graphical user interface for the stocks application



Appointments and Reminders

The appointments and reminders application manages the user's calendar and allows the user to send reminders to other people. It supports time-based requests of various forms, for example, "Remind me to go to the dentist on Monday," "Remind me to see the dentist on August 15^{th}," and "Remind me to see the dentist a week from today." It displays an easily readable schedule, so that the user can recall what is planned for the day. The appointments and reminders application interacts with other applications, such as the shopping application. For example, the request "Remind John to go shopping on Monday" sends a reminder to John, along with the current shopping list, if the user wishes to forward that information. It also supports reminders to the user that are location-aware using GPS, for example, if the user is in the vicinity of a supermarket. The appointments and reminders application is an extension of a calendar service. It links to Microsoft Office Outlook®, and updates scheduled appointments and reminders when in the vicinity of the user's desktop.

Program Manager

The program manager evaluates sentence fragments from a user's request, identifies keywords that determine which application or applications should process the request, reparses the sentence using the grammars for those applications, and forwards the parsed request to the appropriate application. If more than one user is involved, the program manager on one user's mobile device sends messages to the program manager on another user's mobile device, which then handles the request.

The program manager leverages DynaSpeak and a weighted keyword recognition algorithm to break down recognized sentences into application-specific fragments. Those fragments are then processed by the appropriate applications, and are subsequently merged to form the final sentence meaning. This process allows the program manager to handle requests that involve more than one application, for example, "Search for a gas station around Paul Green's house." The parsing of this sentence, using the location grammar, requests a search centered on a location that the location grammar cannot itself provide. The program manager must recognize a keyword from the contacts grammar, parse the sentence using that grammar, and query the contacts application for the address of Paul Green's house. The response to the query is then sent to the location application to obtain the location of the gas station nearest his house.

Graphical User Interface

The graphical user interface (GUI) of thepProgram manager, shown in Figure 1, displays the current running application programs and allows the user to select an application by using voice or keyboard input. The GUI provides buttons that appear gray when an application has not been started and blue after startup. If the user makes a spoken request that requires an application to display a result, the display for that application is topmost and remains topmost until the user issues another request or a timeout occurs. Whenever the GUI is displayed, the user must provide a keyword in a spoken request to wake up the program manager, or click on one of the application-specific buttons on the display.

EVALUATION

Several experiments were performed to collect qualitative and quantitative data to evaluate the prototype system. Although it is difficult to determine a clear boundary between the user interface and the speech recognizer, it is important to evaluate the user interface and the speech recognizer separately, so that the qualitative and quantitative data gathered from the experiments are not mixed, leading to inconclusive results.

Thus, the experiments were designed as a classical "Don't mind the man behind the curtain" study. In this type of study, the user interacts with a system that is identical to the actual system except that the experiment is being controlled by someone other than the user. The man behind the curtain controls what is spoken as responses to the user's requests and changes the current screen to an appropriate graphical response. This method was used, so that the responses to the qualitative questions would not be biased by the accuracy of the speech recognizer.

To evaluate the system quantitatively, the program manager was instrumented with time segment metrics and data were collected for several performance metrics, including:

- Total time a participant took to complete all tasks
- Overhead of the DynaSpeak speech recognizer during live and batch recognition
- Runtime overhead of the program manager without DynaSpeak
- Spoken length of a request vs. processing time

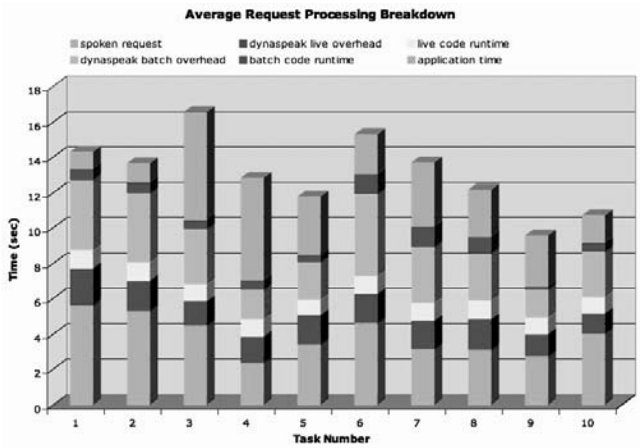
The results are shown in Figure 6. The time segment metrics represent the runtime complexity of the code associated with the speech recognition and processing. The amount of time taken by each segment adds to the delay associated with the user's request. If any of the time segments has a large duration, the user might become irritated. By measuring each segment separately, the bottleneck in the system can be determined.

The speech processing time increases with the size of the grammar. However, by means of a multiphase procedure that uses keywords organized and weighted by application relevance, the grammar size and the speech processing time can be improved. After live recognition, the system provides a keyword-associated request, which it processes for application weights and then reprocesses using an application-specific grammar, possibly more than once with different grammars. This procedure increases both the speed and the accuracy of the speech recognition, by decreasing the size of the grammar size in the initial phase.

An alternative approach (Kondratova, 2004) is to force the user to make repeated requests, possibly from a menu, with responses by which the device asks for the next step or for more information, so that the device arrives at a better understanding of the user's request. Such an approach introduces navigational complexity for the user. Reducing the speech processing time by creating a complex navigational structure is not the best way to improve usability of the system.

The speech recognizer works better for some speakers than for other speakers. The accuracy of the results can be improved by tuning the speech recognition parameters and enabling learning capabilities. However, the developers of DynaSpeak advise against modification of the speech recognition parameters and use of learning until a relatively high success rate is achieved. For appropriately selected users, quite good speech recognition and understanding can be achieved

Figure 6. Processing overhead per task



without using learning capabilities. However, speech recognition accuracy can only improve if voice profiling is combined with learning.

Ambient noise and microphone quality also affect speech recognition accuracy. The internal microphone in the OQO device is of rather poor quality. To ameliorate this problem, a Jabra© Bluetooth headset, was used to provide noise cancellation and reduce the distance between the microphone and the user's mouth. In addition, when the confidence score from DynaSpeak falls below an acceptable threshold, the program manager seeks confirmation from the user or asks for clarification. These mechanisms greatly improve the accuracy of the speech recognizer.

The accuracy of speech recognition is degraded when the grammar contains words that are phonetically similar. During preliminary experiments for the shopping application, we had problems recognizing differences between similar sounding requests like "Add lamb to my shopping list" and "Add ham to my shopping list." These problems arise particularly when users are non-native English speakers or when they have accents. Creating more specific requests can reduce the phonetic similarity, for example, by saying "Add a lamb shank to my shopping list" and "Add a ham hock to my shopping list." However, modifying requests in such a way is undesirable because the requests are then less intuitive and natural.

The location, weather, and stocks applications all use Web Services and require communication over the Internet and, thus, have longer application runtimes than the other Web Services. The location application is written in Java, which runs more slowly than C#. Both the weather application and the stocks application cache data associated with previous requests to take advantage of timing locality. Location requests are different because the caching of maps can involve a large usage of the memory, and users are not inclined to perform the same search twice. Memory is a precious commodity on a handheld device and needs to be conserved; thus, the location application is coded so that it does not cache maps resulting from previous queries.

To evaluate the qualitative aspects of the system, we performed a user study with participants from diverse backgrounds of education, ethnicity, and sex. The user study was completed with 10 individuals performing 10 tasks resulting in 100 request results. The participants were given a questionnaire that assessed their general impressions about the prototype, with the results shown in Table 1.

After analyzing the averaged responses of the participants, we found several trends. The participants' scores are not strongly correlated with speech recognition accuracy. Participant G gave the system a high score, but was one of the two

Table 1. Responses to the questionnaire

Questions	A	B	C	D	E	F	G	H	I	J	Mean
Was it comfortable talking to the device as if it were a human?	3	3	4	3	4	3	4	5	3	5	3.7
Was the GUI aesthetically pleasing?	5	4	4	5	5	5	5	5	5	4	4.7
Were the request responses appropriate and easy to understand?	3	3	5	5	5	4	4	5	4	4	4.2
Were the spoken responses relevant to your requests?	5	4	5	5	5	3	4	5	4	5	4.5
Was the system easy to use?	4	5	3	4	4	5	4	5	4	5	4.3
Do you think the services would be helpful in your daily life?	4	4	4	5	4	4	4	4	4	5	4.2
Would you recommend a system like this to your friends?	3	3	4	5	4	4	5	5	4	5	4.2
Would you buy the software if it were available for your phone?	3	2	4	5	3	3	5	5	3	5	3.8

participants who encountered the most speech recognition problems. Participant B gave the system a low score despite good speech recognition.

The participants agreed that speaking to a mobile handheld device as if it were a human is not comfortable. It is difficult to get used to interacting with a computer that can understand tasks that would be commonplace for humans. The participants were relatively pleased with the GUI interface design and felt the system is relatively easy to use. However, the ease-of-use metric needs to be taken lightly. Ease of use can be assessed more concretely by measuring the number of times a user must repeat a command.

The scores for response appropriateness and relevance are high, indicating that the spoken responses of the applications were well crafted. The scores related to recommending the service to friends and daily life helpfulness are relatively high, from which one might infer that the participants would purchase a device providing the speech-enabled applications. However, this conclusion is not necessarily justified. The participants were not enthusiastic about having to pay for such a device or for such services. However, most participants in the study were quite pleased with the prototype system and found the user interface helpful and easy to use.

FUTURE TRENDS

Integration of multiple applications, and multiple grammars, is not too difficult for a small number of applications that have been designed and programmed to work together, as in our prototype. However, future systems will need to support tens or hundreds of applications, many of which will be designed and programmed independently. Integration of those applications and their grammars will be a challenge.

Currently, speech-enabled applications typically use short commands from the human that are translated into navigational or functional operations. More appropriate is speech recognition technology that supports a more natural, conversational style similar to what humans use to communicate with each other (McTear, 2002).

A mobile device that listens to its owner continuously can provide additional services, such as populating the user's calendar. For example, when a user agrees to an appointment during a conversation with another person, the mobile device might recognize and automatically record the appointment, possibly confirming the appointment later with its user. Similarly, the mobile device might note that the user habitually goes to lunch with the gang at noon on Mondays, or that the user leaves work promptly at 5pm on Fridays. With existing calendar systems, the user often does not record appointments and other commitments, because it is too much bother using the human interfaces of those systems, greatly reducing the value of the calendar.

A useful capability of speech recognition systems for mobile devices is being able to recognize intonation and emotional overtones. "The bus leaves at 6" is, overtly, a simple declaration, but appropriate intonation might convert that declaration into a question or an expression of disapproval. Existing speech recognition systems do not yet recognize and exploit intonation. Similarly, the ability to recognize emotional overtones of impatience, uncertainty, surprise, pleasure, anger, and so forth, is a valuable capability that existing speech recognition systems do not yet provide.

Speech recognition requires a relatively powerful processor. Typical cell phones contain a powerful digital signal processor (DSP) chip and a much less powerful control processor. The control processor operates continuously to maintain communication with the cellular base stations. The DSP processor uses a lot of power and imposes a significant drain on the battery and, thus, analyzes and encodes speech only during calls. The DSP processor is capable of the processing required for speech recognition, although it might need more memory.

For mobile devices, battery life is a problem, particularly when speech recognition or application software requires a powerful processor. The limit of 2 hours of talk time for a cell phone is caused at least as much by the power drain of the DSP processor as by the power needed for wireless transmission. The DSP processor might be needed for speech processing for more than 2 hours per day. There are several possible solutions to this problem, namely, larger batteries, alcohol fuel cells, and DSP processors with higher speeds, reduced power consumption, and better power management.

Background noise remains a problem for speech recognition systems for mobile devices, particularly in noisy environments. The quality of the microphone, and the use of a headset to decrease the distance between the microphone and the speaker's mouth, can improve speech recognition accuracy.

CONCLUSION

The use of voice input and output, in addition to text and graphics and other kinds of audio, video, and tactile interfaces, provides substantial benefits for the users of mobile devices. Such multimodal interfaces allow individuals to access information, applications, and services from their mobile devices more easily. A user no longer has to put up with the annoyances of a 3-inch keyboard, nested menus, or handwriting recognition, nor does the user need to have a tethered desktop or server computer in order to access information, applications, and services. Providing multiple ways in which the users can interact with the applications on mobile devices brings a new level of convenience to the users of those devices.

REFERENCES

AT&T. (2006). *Natural voices.* Retrieved from http://www.natural voices.att.com/products/

Booth, D., Hass, H., McCabe, F., Newcomer, E., Champion, M., Ferris, C., & Orchard, D. (2004). *Web services architecture*. Retrieved from http://www.w.3.org/Tr/WS-arch

Deligne, S., Dharanipragada, S., Gopinath, R., Maison, B., Olsen, P., & Printz, H. (2002). A robust high accuracy speech recognition system for mobile applications. *IEEE Transactions on Speech and Audio Processing*, *10*(8), 551-561.

Deng, L., & Huang, X. (2004). Challenges in adopting speech recognition. *Communications of the ACM*, *47*(1), 69-75.

Deng, L., Wang, K., Acero, A., Hon, H., Droppo, J., Boulis, C., Wang, Y., Jacoby, D., Mahajan, M., Chelba, C., & Huang, X. D. (2002). Distributed speech processing in MiPad's multimodal user interface. *IEEE Transactions on Speech and Audio Processing*, *10*(8), 605-619.

Frost, R. A. (2005). Call for a public-domain SpeechWeb. *Communications of the ACM, 48*(11), 45-49.

Google. (2006). *Google Maps API.* Retrieved from http://www.google.com/apis/maps

Grasso, M. A., Ebert, D. S., & Finin, T. W. (1998). The integrality of speech in multi-modal interfaces. *ACM Transactions on Computer-Human Interaction, 5*(4), 303-325.

Hjelm, J. (2000). *Research applications in the mobile environment. Wireless information service.* New York, NY: John Wiley & Sons.

Holtzblatt, K. (2005). Designing for the mobile device: Experiences, challenges, and methods. *Communications of the ACM, 48*(7), 33-35.

Huang, X., Acero, A., Chelba, C., Deng, L., Droppo, J., Duchene, D., Goodman, J., Hon, H., Jacoby, D., Jiang, L., Loynd, R., Mahajan, J., Mau, P., Meredith, S., Mughal, S., Neto, S., Plumpe, M., Stery, K., Venolia, G., Wang, K., & Wang, Y. (2001). MiPad: A multimodal interaction prototype. In *Proceedings of the International Conference on Acoustics, Speech and Signal Processing*, *1*, 9-12.

Knight, S., Gorrell, G., Rayner, M., Milward, D., Koeling, R., & Lewin, I. (2001). Comparing grammar-based and robust approaches to speech understanding: A case study. In *Proceedings of Eurospeech 2001, Seventh European Conference of Speech Communication and Technology* (pp. 1779-1782). Aalborg, Denmark.

Kondratova, I. (2004, August). Speech-enabled mobile field applications. In *Proceedings of the IASTED International Conference on Internet and Multimedia Systems*, Hawaii.

Kondratova, I. (2005, July). Speech-enabled handheld computing for fieldwork. In *Proceedings of the International Conference on Computing in Civil Engineering*, Cancun, Mexico.

Le Bodic, G. (2002). *Mobile messaging, SMS, EMS and MMS*. John Wiley & Sons.

McTear, M. (2002). Spoken dialogue technology: Enabling the conversational user interface. *ACM Computing Surveys, 34*(1), 90-169.

National Oceanic and Atmospheric Administration (NOAA). (2006). *National Weather Service*. Retrieved from http://www.weather.gov/xml/

OQO. (2006). *The OQO personal computer*. Retrieved from http://www.oqo.com

Oviatt, S., & Cohen, P. (2000). Multi-modal interfaces that process what comes naturally. *Communications of the ACM, 43*(3), 45-53.

Rabiner, L., & Juang, B. H. (1993). *Fundamentals of speech recognition*. Upper Saddle River, NJ: Prentice Hall.

Rao, B., & Minakakis, L. (2003). Evolution of mobile location-based services. *Communications of the ACM, 46*(12), 61-65.

Sarker, S., & Wells, J. D. (2003). Understanding mobile handheld device use and adoption. *Communications of the ACM, 46*(12), 35-40.

SRI (2006). *DynaSpeak*. Retrieved from http://www.speechatsri.com/products/sdk.shtml

Srinivasan, S., & Brown, E. (2002). Is speech recognition becoming mainstream?. *Computer Magazine*, (April), 38-41.

U.S. Census Bureau. (2006). *Precision of GPS*. Retrieved from http://www.census.gov/procur/www/fdca/library/mcd/7-29%20 MCD_WG_hardware_subteam_report.pdf

Varga, I., Aalburg, S., Andrassy, B., Astrov, S., Bauer, J. G., Beaugeant, C., Geissler, C., & Hoge, H. (2002). ASR in mobile phones—An industrial approach. *IEEE Transactions on Speech and Audio Processing, 10*(8), 562-569.

Yahoo! LocalSearch. (2006a). Retrieved from http://www.local.yahooapis.com/LocalSearchService/V3/localSearch

Yahoo! Finance. (2006b). Retrieved from http://finance.yahoo.com/rssindex

Zhang, W., He, Y., Chow, R., Yang, R., & Su, Y. (2000, June). The study on distributed speech recognition system. In *Proceedings of the IEEE International Conference on Acoustical Speech and Signal Processing* (pp. 1431–1434), Istanbul, Turkey.

KEY TERMS

Global Positioning System (GPS): A system that is used to obtain geographical coordinates, which includes a GPS satellite and a GPS receiver.

Hidden Markov Model (HMM): A technique, based on a finite state machine that associates probabilities with phonemes, and pairs of phonemes, that is used in speech recognition systems, to determine the likelihood of an expression spoken by a user of that system.

Location Aware: An application that is based on a particular physical location, as given by geographical coordinates, physical address, zip code, and so forth, that determines the output of the application.

Mobile Device: For the purposes of this chapter, a handheld device, such as a cell phone or personal digital assistant (PDA), that has an embedded computer and that the user can carry around.

Multimodal Interface: The integration of textual, graphical, video, tactile, speech, and other audio interfaces through the use of mouse, stylus, fingers, keyboard, display, camera, microphone, and/or GPS.

Speech Recognition: The process of interpreting human speech for transcription or as a method of interacting with a computer or a mobile device, using a source of speech input, such as a microphone.

Speech Synthesis: The artificial production of human speech. Speech synthesis technology is also called text-to-speech technology in reference to its ability to convert text into speech.

Web Service: A software application identified by a Uniform Resource Indicator (URI) that is defined, described, and discovered using the eXtensible Markup Language (XML) and that supports direct interactions with other software applications using XML-based messages via an Internet protocol.

Chapter XXVIII
Speech-Centric Multimodal User Interface Design in Mobile Technology

Dong Yu
Microsoft Research, USA

Li Deng
Microsoft Research, USA

ABSTRACT

Multimodal user interface (MUI) allows users to interact with a computer system through multiple human-computer communication channels or modalities. Users have the freedom to choose one or more modalities at the same time. MUI is especially important in mobile devices due to the limited display and keyboard size. In this chapter, we provide a survey of the MUI design in mobile technology with a speech-centric view based on our research and experience in this area (e.g., MapPointS and MiPad). In the context of several carefully chosen case studies, we discuss the main issues related to the speech-centric MUI in mobile devices, current solutions, and future directions.

INTRODUCTION

In recent years, we have seen steady growth in the adoption of mobile devices in people's daily lives as these devices become smaller, cheaper, more powerful, and more energy-efficient. However, mobile devices inevitably have a small display area, a tiny keyboard, a stylus, a low speed (usually less than 400 million instructions per second) central processing unit (CPU), and a small amount (usually less than 64MB) of dynamic random-access memory. Added to these limitations is the fact that mobile devices are often used in many different environments, such as dark and/or noisy surroundings, private offices, and meeting rooms. On these devices, the traditional *graphical user* *interface* (GUI)-centric design becomes far less effective than desired. More efficient and easy-to-use user interfaces are in urgent need. The *multimodal user interface* (MUI), which allows users to interact with a computer system through multiple channels such as speech, pen, display, and keyboard, is a promising user interface in mobile devices.

Multimodal interaction is widely observed in human-human communications where senses such as sight, sound, touch, smell, and taste are used. The research on multimodal human-computer interaction, however, became active only after Bolt (1980) proposed his original concept of "Put That There." Since then, a great amount of research has been carried out in this area (Bregler, Manke, Hild, & Waibel 1993; Codella, Jalili, Koved, Lewis, Ling, Lipscomb, et al., 1992; Cohen, Dalrymple, Moran, Pereira, Sullivan, Gargan, et al., 1989; Cohen, Johnston, McGee, Oviatt, Pittman, Smith, et al., 1997; Deng & Yu, 2005; Fukumoto, Suenga, & Mase, 1994; Hsu, Mahajan, & Acero 2005; Huang, Acero, Chelba, Deng, Droppo, Duchene, et al., 2001; Neal & Shapiro, 1991; Pavlovic, Berry, & Huang, 1997; Pavlovic & Huang, 1998; Vo, Houghton, Yang, Bub, Meier, Waibel, et al., 1995; Vo & Wood, 1996; Wang, 1995). Importantly, the body of this research work pointed out that MUIs can support flexible, efficient, and powerful human-computer interaction.

With an MUI, users can communicate with a system through many different input devices such as keyboard, stylus, and microphone, and output devices such as graphical display and speakers. MUI is superior to any single modality where users can communicate with a system through only one channel. Note that using an MUI does not mean users need to communicate with the system always through multiple communication channels simultaneously. Instead, it means that users have freedom to choose one or several modalities when communicating with the system, and they can switch modalities at any time without interrupting the interaction. These characteristics make the MUI easier to learn and use, and is preferred by users in many applications that we will describe later in this chapter.

MUI is especially effective and important in mobile devices for several reasons. First, each modality has its strengths and weaknesses. For this reason, single modality does not permit the user to interact with the system effectively across all tasks and environments. For example, speech UI provides a hands-free, eyes-free, and efficient way for users to input descriptive information or to issue commands. This is very valuable when in motion or in natural field settings. Nevertheless, the performance of speech UI decreases dramatically under noisy conditions. In addition, speech UI is not suitable when privacy and social condition (e.g., in a meeting) is a concern. Pen input, on the other hand, allows users to interact with the system silently, and is acceptable in public settings and under extreme noise (Gong, 1995; Holzman, 1999). Pen input is also the preferred way for entering digits, gestures, abbreviations, symbols, signatures, and graphic content (Oviatt & Olsen, 1994; Suhm, 1998). However, it is impossible for the user to use pen input if he/she is handicapped or under "temporary disability" (e.g., when driving). MUI, on the other hand, allows users to shift between modalities as environmental conditions change (Holzman, 1999), and hence, can cover a wider range of changing environments than single-modal user interfaces.

Second, different modalities can compensate for each other's limitations and thus provide users with more desirable experience (Deng & Yu, 2005; Oviatt, Bernard, & Levow, 1999; Oviatt & vanGent, 1996; Suhm, 1998). For example, the accuracy of a resource-constrained, midsized vocabulary speech recognizer is low given the current speech technology. However, if the speech recognizer is used together with a predictive T9 (text on 9 keys) keyboard, users can greatly increase the text input throughput compared with using the speech modality or T9 keyboard alone (Hsu et al., 2005). The gain is obtained from the mutual disambiguation effect, where each error-prone modality provides partial information to aid in the interpretation of other modalities. Another reason for the improved user experience is users' active error avoidance, where users tend to select the input modality that they judge to be less error prone for a particular task and environment (Oviatt & vanGent, 1996), and tend to switch modalities to recover from system errors (Oviatt et al., 1999). Mutual compensation is very important for mobile devices because the ability of every single modality in the devices is extremely limited (e.g., a limited display and keyboard size, and limited speech recognition accuracy).

Despite the importance of MUI in mobile devices, designing effective MUIs is far from trivial. Many MUIs in mobile devices are *speech centric*, where speech is the central and main modality. In this chapter, we will focus on main issues on the design of effective speech centric MUIs in mobile devices based on our research and experience in developing MapPointS (Deng & Yu, 2005) and MiPad (Deng, Wang, Acero, Hon, Droppo, Boulis, et al., 2002; Huang, Acero, Chelba, Deng, Droppo, Duchene, et al., 2001). In Section 2, we describe a generic MUI architecture in mobile setting that consists of various recognizers for different input modalities, semantic parsers, a discourse manager, and a response manager. In Section 3, we discuss special considerations related to speech modality. In particular, we discuss the approaches to overcoming resource limitations on mobile devices, noise robust speech front-ends, noise robust modality switching interfaces, and context-aware language model. In section 4, we introduce the issues related to robust natural language understanding including construction of robust grammars. We discuss the problem of modality fusion, including modality-neutral semantic representation, unification approach, and modality integration, in Section 5. We discuss possible future directions and conclude this chapter in Section 6.

A GENERIC MUI ARCHITECTURE

The ultimate goal of an MUI is to fulfill the needs and requirements of the users. This principle is one of many emphasized in *user-centered design* (Gould & Lewis, 1985, Norman & Draper, 1986). According to the user-centered design principle, the acceptability of an MUI can be judged using three main attributes (Dybkjaer & Bernsen, 2001; Hone & Graham, 2001; Nielsen, 1993): effectiveness, efficiency, and learnability. The *effectiveness* assesses whether users can complete the tasks and achieve the goals with the predefined degree of perceived accuracy. It is usually measured on the targeted user population, over a specified range of tasks and environments. The *efficiency* judges how much effort (cognitive demand, fatigue, stress, frustration, discomfort, and so on) and resources (time) are needed for users to perform specific tasks. It is usually measured with the total time (including time for error corrections) taken to complete a task. The *learnability* measures whether users can easily discover the system's functionality and quickly learn to use the system.

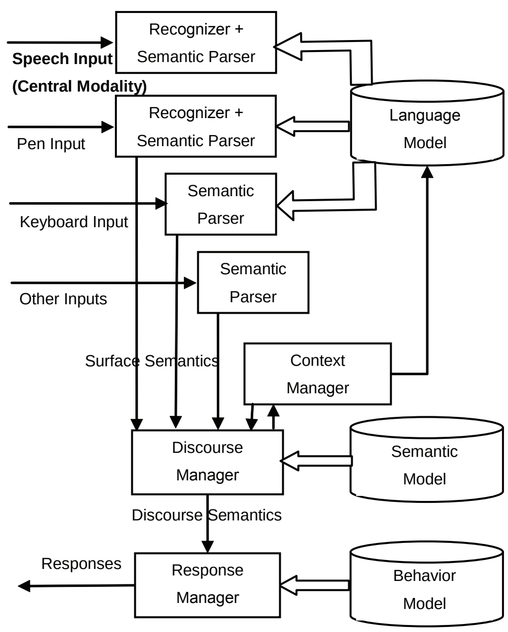
Figure 1 depicts a typical speech-centric MUI architecture that is aimed to achieve a high level of effectiveness, efficiency, and learnability. As shown in the figure, users can communicate with the system through speech, keyboard, and other modalities such as pen and camera. Modality fusion usually is the center of an MUI system. There are two typical ways of fusing information from different input modalities, namely, early fusion and late fusion. With the *early fusion*, signals are integrated at the *feature* level and hence, the recognition process in one modality would affect that in another modality (Bregler et al., 1993, Pavlovic et al., 1997; Pavlovic & Huang, 1998; Vo et al., 1995,). Early fusion is suitable for highly coupled modalities such as speech and lip movements (Rubin, Vatikiotis-Bateson, & Benoit, 1998; Stork & Hennecke, 1995). However, early fusion can greatly increase the modeling complexity and computational intensity due to its nature of intermodality influence in the recognition phase. With the *late fusion*, information is integrated at the *semantic* level. The benefit of late fusion is its isolation of input modalities from the rest of the system. In other words, individual recognizers trained using unimodal data can be directly plugged into the system without affecting the rest of the system. This feature makes the late fusion easier to scale up to more modalities in the future than the early fusion. The architecture shown in Figure 1 utilizes the late fusion approach that has been widely adopted, for example, by a variety of systems including Put-That-There (Bolt, 1980), MapPointS (Deng & Yu, 2005), MiPad (Huang et al., 2001), ShopTalk (Cohen, et al., 1989), QuickSet (Cohen, Johnston, McGee, Oviatt, Pittman, Smith, et al., 1997), CUBRICON (Neal & Shapiro, 1991), Virtual World (Codella, Jalili, Koved, Lewis, Ling, Lipscomb, et al., 1992), Finger-Pointer (Fukumoto et al., 1994), VisualMan (Wang, 1995), and Jeanie (Vo & Wood, 1996).

In the late-fusion approach depicted in Figure 1, the input signals received by the system are first processed by *semantic parsers* associated with the corresponding modality into the *surface semantics* representation. Note that although each modality has its own semantic parser, the resulting surface semantics are represented in a common semantic representation and is thus independent of the modality. The surface semantics from all the input modalities are then fused by the *discourse manager* component into the *discourse semantics* representation (more discussions on this issue in Section 4). In order to generate discourse semantics, the discourse manager uses the semantic modal and interacts with the context manager to utilize and update such information as dialog context, domain knowledge, user's information, and user's usage history. The updated context information can be used to adapt the language model, which can improve speech recognition accuracy and enhance the quality of semantic parsers for the next user-computer interaction.

The discourse semantics, which is the output of the discourse manager, is then fed into the *response manager* to communicate back to the user. The response manager synthesizes the proper responses, based on the discourse semantics and the capabilities of the user interface, and plays the response back to the user. In this process, behavior model provides rules to carry out the required actions. The combination of discourse manager and response manager is usually referred to as the dialog manager.

Note that the components shown in Figure 1 may reside on the mobile devices, or distributed on other servers in real implementations. In addition, many MUI systems use an agent-based software solution in which a facility or hub is used to pass information to and from different components (or agents) (Kumar & Cohen, 2000; Schwartz, 1993).

Figure 1. A typical speech-centric MUI architecture and its components



Many best practices and design principles have been developed for the speech-centric MUI design in the past decades (Becker, 2001; Dybkjaer & Bernsen, 2001; Ravden & Johnson, 1989; Reeves, Lai, J., Larson, J.A., Oviatt, S., Balaji, T.S., Buisine, et al. 2004), which we summarize next.

First, the system should explicitly inform the user about its state through appropriate feedback within a reasonable amount of time, so as to avoid state errors, that is, the user's perceived state is different from the system's perceived state. The feedback can be in different modalities, but must be clear and accurate. If speech feedback is used, recorded speech is usually preferred over the synthesized speech, due to its higher degree of naturalness. Note that the recorded speech usually takes a larger amount of resources than the synthesized speech. Since the memory and storage available in mobile devices is very limited, designers should strike a balance between the use of synthesized speech and of recorded speech. The system should follow real-world conventions, and use the words, phrases, and concepts that are familiar to the users. The system should also ensure that the output modalities be well synchronized temporally. For example, the spoken directions should be synchronized with the map display.

Second, the system should provide sufficient flexibility so that users can select the modalities that are best for the task under the specific environments. For example, the user should be able to switch to a nonspeech modality when inputting sensitive information such as personal identification numbers and passwords. A good MUI design should also allow users to exit from an unwanted state via commands that are global to the system, instead of having to go through an extended dialog. The system should provide enough information (e.g., through prompts) to guide novice users to use the system, yet at the same time allow barge-ins and accelerators for the expert users to reduce the overall task completion time.

Third, the system should be designed to allow easy correction of errors. For example, the system should provide context sensitive, concise, and effective help. Other approaches include integrating complementary modalities to improve overall robustness during multimodal fusion; allowing users to select a less error-prone modality for a given lexical content, permitting users to switch to a different modality when error happens; and incorporating modalities capable of conveying rich semantic information.

Fourth, the system's behavior should be consistent internally and with users' previous experiences. For example, a similar dialog flow should be followed and the same terms should be used to fulfill the same task. Users should not have to wonder whether the same words and actions have different meaning under different context.

Fifth, the system should not present more information than necessary. For example, dialogues should not contain irrelevant or rarely needed information, and the prompts should be concise.

While the best practices summarized are common to all speech-centric MUIs, some special attention needs to be paid to speech modality and multimodality fusion due to the great variations of mobile device usage environments. We address these special considerations next.

SPECIAL CONSIDERATIONS FOR SPEECH MODALITY

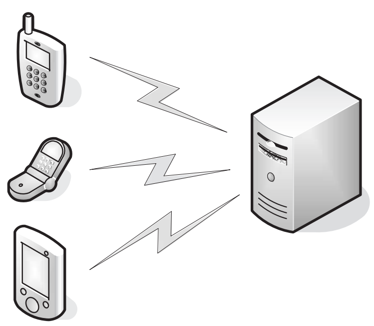
There are two main challenges for the use of speech modality on mobile devices. First, the resources on mobile devices, in particular, CPU speed, memory, and communication bandwidth, are very limited. Second, speech recognition accuracy degrades substantially in realistic noisy environments, where there are abrupt changes in noise, or variable phase-in phase-out sources of noise as the user moves. For example, the recognition accuracy may drop 30-50% inside a vehicle and cafeteria from that in a quiet environment (Das, Bakis, Nadas, Nahamoo, & Picheny, 1993; Lockwood & Boudy, 1992). Since the mobile devices will be used in these real-field settings without a close-talk microphone, robustness to acoustic environment, that is, immunity to noise and channel distortion, is one of the most important aspects to consider when designing speech-centric MUIs on mobile devices. Speech recognition accuracy and robustness can usually be improved with a noise-robust speech front-end, a noise-robust modality-switching interface, and a context aware language model.

Resource Constrained Speech Recognition

Speech recognition on mobile devices is typically carried out with two options: the *distributed recognition* (Deng et al., 2002) where the recognition happens at a remote server (Figure 2) and the *local recognition* (Deligne, Dharanipragada, Gopinath, Maison, Olsen, & Printz, 2002; Varga, Aalburg, Andrassy, Astrov, Bauer, Beaugeant, et al., 2002) where the recognition is carried out completely on the mobile device. The distributed recognition can take advantage of the power of the remote server to achieve a fast and accurate recognition, while the local recognition can eliminate the requirement of the device to have a fast data connection.

In the distributed architecture, the main consideration is the latency required to send data to and from the server. The latency is typically determined by the communication bandwidth and the amount of data sent. To reduce the latency, a typical approach is to use a standard codec on the device to transmit the speech to the server where the coded speech is subsequently decompressed and recognized (as depicted in Figure 3). However, since speech recognizers only need some features of the speech signal (e.g., Mel-cepstrum), an alternative approach is to put the speech front end on the mobile device and transmit only speech features to the server (Deng et al. 2002), as shown in Figure 4. Transmitting speech features can further save bandwidth because the size of the features is typically much less than that of the compressed audio signals.

Figure 2. Illustration of distributed speech recognition where the actual recognition happens at the server (e.g., PC)



Besides the advantage of using the computing power at the server to improve speech recognition accuracy, there are other benefits of using server-side recognition. One such benefit is its better maintainability compared to the local recognition approach because updating software on the server is much easier and more cost effective than updating software on millions of mobile devices. It, however, does require the recognizer on the server to be front end or codec agnostic in order to materialize this benefit. In other words, the recognizer should make no assumptions on the structure and processing of the front end (Deng et al., 2002). Another benefit of using distributed recognition is the possibility for the server to personalize the acoustic model, language model, and understanding model all at the server, saving the precious CPU and memory on mobile devices. In the past, distributed recognition is unquestionably the dominant approach due to the low CPU speed and small amount of memory available on the mobile devices. Nowadays, although the CPU speed and memory size are increasing dramatically, distributed recognition is still the prevailing approach over local recognition due to the advantages discussed previously.

The major issue of the local recognition architecture is the low recognition speed and accuracy due to the slow CPU speed and low memory available on mobile devices. Speech recognizers running on mobile devices need to be specially designed (Deligne et al., 2002, Li, Malkin, & Bilmes, 2006; Varga, Aalburg, Andrassy, Astrov, Bauer, Beaugeant, 2002) to fit the requirement since speech recognizers designed for the desktop or telephony systems cannot be directly deployed to mobile devices. The greatest benefit of using the local recognition approach is its independency of the network connection and the server and

Figure 3. Distributed speech recognition architecture: speech input is encoded and sent to the server. Speech feature extraction happens at the server side

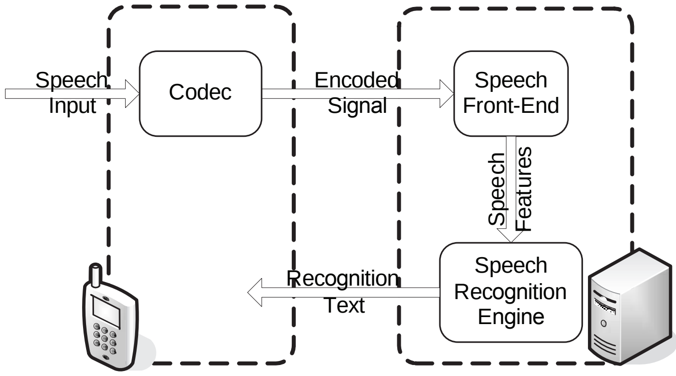
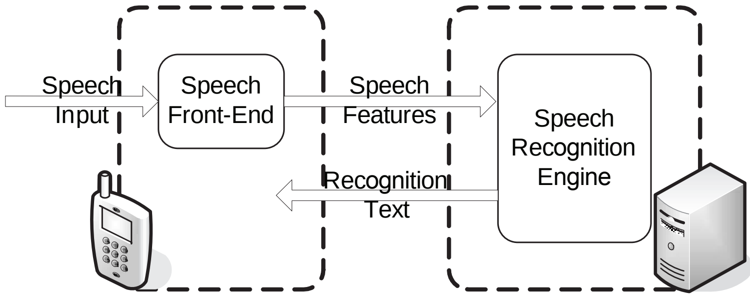


Figure 4. Distributed speech recognition architecture alternative: the speech feature extraction happens on the mobile devices. Only the features are sent to the server



hence, can be used everywhere under any conditions. Given the consistent improvement of the CPU speed and memory on the mobile device hardware, in the future, the local recognition approach is expected to become more and more popular for simple tasks such as name dialing and media playing.

Noise Robust Speech Front End

Noise robustness is one of the most important requirements for speech-centric MUI on mobile devices. It has attracted substantial attention in the past several years. Many algorithms have been proposed to deal with nonstationary noises. A popular one is an advanced feature extraction algorithm (jointly developed by Motorola Labs, France Telecom and Alcatel) that was selected in February of 2002 as a standard in distributed speech recognition by the European telecommunications standards institute. The algorithm defines the extraction and compression of the features from speech that is performed on a local, terminal device, for example, a mobile phone. These features are then sent over a data link to a remote “back-end processor” that recognizes the words spoken. The major components of this algorithm are noise reduction, waveform processing, cepstrum calculation, blind equalization, and voice-activity detection. The noise reduction component makes use of two-stage Wiener filtering (Macho, Mauuary, Noé, Cheng, Ealey, Jouvet, et al., 2002).

The stereo-based piecewise linear compensation for environments (SPLICE), which has been used in the MiPad system (Deng et al., 2002), is another effective algorithm for noise robust speech feature extraction. SPLICE is a cepstrum enhancement algorithm dealing with additive noise, channel distortion, or a combination of the two. It is a dynamic, frame-based, bias-removal algorithm with no explicit assumptions made on the nature of the noise model. In SPLICE, the noise characteristics are embedded in the piecewise linear mapping between the "stereo'' clean and distorted speech cepstral vectors. SPLICE has a potential to handle a wide range of distortions, including nonstationary distortion, joint additive and convolutional distortion, and nonlinear distortion (in time-domain), because SPLICE can accurately estimate the correction vectors without the need for an explicit noise model.

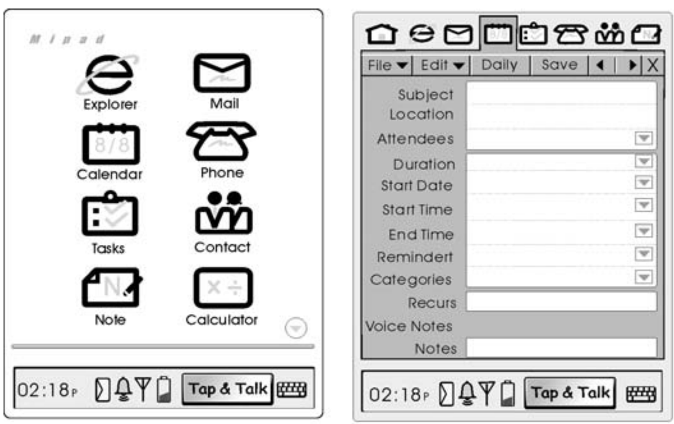
Modality Switching

One of the problems in speech recognition under noisy environment is modality switching. If the speech recognition engine is always on, noises and by-talks may be misrecognized as a legitimate user input and hence, can erroneously trigger commands.

A widely used modality switching approach is called "push to talk," where the user presses a button to turn on the speech recognizer, and releases the button to turn off the recognizer. Another approach is called "tap & talk" (Deng et al., 2002; Huang, Acero, A., Chelba, C., Deng, L., Duchene, D., Goodman, et al., 2000, Huang et al., 2001), where the user provides inputs by tapping the "tap & talk" field and then talking to it. Alternatively, the user can select the tap & talk field by using the roller to navigate and holding it down while speaking. Tap & talk can be considered as a combination of push-to-talk control and indication of where the recognized text should go. Both the push-to-talk and tap & talk avoid the speech detection problem that is critical to the noisy environment under which the mobile devices are typically deployed.

Figure 5 shows an example of the tap & talk interface used in the MiPad (Deng et al., 2002). If the user wants to provide the attendee information for a meeting scheduling task, he/she taps the "attendees" field in the calendar card. When that happens, the MUI will constrain both the language model and the semantic model based on the information on the potential attendees. This can significantly improve the accuracy and the throughput. Note that tap & talk functions as a user-initiative dialog-state specification. With tap & talk, there is no need for the mobile devices to include any special mechanism to handle spoken dialog focus and digression.

Figure 5. An example of the Tap & Talk interface (Deng et al., 2002, © 2002 IEEE)



Context-Aware Language Model

Here, *context* refers to any information that can be used to characterize the situation related to human-computer interaction. It typically includes the surrounding environment (e.g., location and noise condition), the user (e.g., age and gender, preferences, past interaction experiences, and the interaction history in the current session), and the devices (e.g., remaining battery life, available memory, screen-size, screen-contrast, and speaker volume). Although context-awareness can be beneficial to all components in an MUI, it is especially important for improving speech recognition accuracy under noisy environments.

Context information can be utilized in many different ways in speech modality. One particular approach is to construct the language model based on the context. For example, the tap & talk approach (Deng et al., 2002) customizes the language model depending on the field the user is pointing to, as mentioned in section 3.3.

Language model can also be customized, based on the user information and the dialog state. For example, if the system is expecting the recipient information, the language model can include only the names in the global address book. If the user information is also used, the language model can also include user's contact list and people who have exchanged e-mails with the user in the past. An even more effective language model would weight different names differently, depending on the frequencies the user exchanged e-mail with the person, and the recentness of the interaction (Yu, Wang, Mahajan, Mau, & Acero, 2003). Another example of constructing the language model based on the context and user information is described in the speech enabled MapPoint (Deng & Yu, 2005). Without context information, the speech recognizer needs to load all location names and business names in the North America. This is definitely beyond the ability of most state-of-the-art speech recognizers. However, if the user's location information and/or the interaction history are known, the system can load only the location names and business names around the user's current location, and weight all the names based on the popularity of the names as well as the user's interaction history.

A more advanced context-aware language model construction technique is discussed by Wang (2004). This detection-based technique is used in the second generation of the MiPad (Wang, 2004). The basic idea of this approach is to detect the context cues from the user's partial utterances sequentially, and adjust the language model dynamically for the next part of the utterances. This approach has achieved excellent user experience.

LANGUAGE UNDERSTANDING

Good speech recognition accuracy does not always translate to good understanding of users' intents, as indicated by Wang, Acero, and Chelba (2003). A robust language-understanding model is needed to obtain good user experience for speech-centric MUI applications, especially since speech recognition errors will affect the understanding.

The first issue to address in language understanding is constructing the semantic grammar. Since the importance of each word to the understanding is different, the words need to be treated differently. A typical approach is to introduce a specific type of nonterminals called semantic classes to describe the schema of an application (Wang, 2001; Yu, Ju, Wang, & Acero, 2006). The semantic classes define the concepts embedded in the linguistic structures, which are usually modeled with probabilistic context-free grammars. The advantage of introducing the semantic classes is to make the linguistic realization of semantic concepts independent of the semantic concepts themselves. Once the semantic classes are defined, a robust linguistic grammar can be built using the approaches similar to the one described by Yu, et al. (2006).

The transformation from the recognized text to the semantic representation is usually done using a semantic parser. For example, in MiPad, this transformation is done using a robust chart parser (Wang, 2001). In this parser, "the robustness to ungrammaticality and noise can be attributed to its ability of skipping minimum unparsable segments in the input. The algorithm uses dotted rules, which are standard context free grammar rules in Backus Naur form plus a dot in front of a right-hand-side symbol. The dot separates the symbols that already have matched with the input words from the symbols that are yet to be matched." (Wang, 2001, pp. 1556) Since the language models used in MiPad are dynamically generated based on the current user information and the tap & talk field, the parser used in MiPad supports dynamic grammars. Given that some part of the user's utterances is in the free-style form (e.g., the topic of a meeting to be scheduled), they are modeled as dictation grammar rules. Since speech recognition is not perfect, the MiPad robust parser takes into account the N-best list, together with the associated confidence scores returned from the speech recognition engine, and combines the speech recognition score with the parsing score to obtain the best parsing result. More recent progress includes using maximum entropy models to classify the tasks and to disambiguate the meaning of the slots in the recognition result.

MODALITY FUSION

One strong advantage of using MUIs is the improved accuracy and throughput through modality integration. There are typically two fusion approaches: early fusion and late fusion. Given that late fusion has many superior properties over the early one, as discussed in Section 2, it will be the focus of our discussion in this section. There are two tasks in the late fusion: Process and convert the input signals into a common surface semantic representation using the semantic parsers (one specific to each modality), and fuse the surface semantics into discourse semantics using the discourse manager.

Semantic Representation and Unification

The semantic fusion operation requires a meaning representation framework that is common among modalities, and a well-defined operation for combining partial meanings.

Many semantic representation formats have been proposed in the past. For example, in Bolt's (1980) pioneering paper, only very limited modality fusion is required and hence, a simple semantic representation was used. In the past decade, researchers (Cheyer & Julia, 1995; Pavlovic & Huang, 1998; Shaikh, Juth, Medl, Marsic, Kulikowski, & Flanagan, 1997; Vo & Wood, 1996) have converged to using a data structure called typed *feature structures* (Kay, 1979) to represent meanings. Typed feature structure can be considered as an extended, recursive version of attribute-value-type data structures, where a value can, in turn, be a feature structure. It extends *frames* (Minsky, 1975) that represent objects and relations as nested sets of attribute/value pairs, by using shared variables to indicate common substructures. A typed feature structure indicates the kind of entity it represents with a type, and the values with an associated collection of feature-value or attribute-value pairs. In the typed feature structure, a value may be nil, a variable, an atom, or another typed-feature structure.

The primary operation on typed feature structure is *unification*. "*Typed-feature-structure unification* is an operation that determines the consistency of two representational structures and, if they are consistent, combines them into a single result." (Oviatt, Cohen, Wu, Vergo, Duncan, Suhm, et. al., 2000, online version pp. 21) Unification can combine complementary input from different modalities and rule out contradictory input (Johnston, 1998).

Note that users' multimodal inputs may involve *sequentially integrated* or *simultaneously delivered* signal fragments. In other words, temporal relationships between different input channels are very important. To fuse modalities, we need to first determine whether two input fragments are related. In most of the systems reported, this is achieved by considering all input contents that lie within a predefined time window. To do this, all input fragments need to be time stamped as soon as they are generated to remove the errors due to transit delays.

For example, the speech input "Show me the restaurants around here." might have a gesture-input accompanying it either "before," "during," or "after" the actual utterance, and all these three possibilities should provide the same result. Usually the term "before" represents a timeframe of up to several minutes, "during" represents a timeframe of 4 to 5 seconds, and "after" represents a timeframe of 500ms to 750ms. If these values are too small, many multimodal inputs will be considered as unimodal inputs and will not be integrated. If the values are too large the chances of an old or invalid user input are likely being accepted as part of a valid multimodal input.

To determine whether two input fragments should be treated as parts of a multimodal construction or separate unimodal commands, knowledge gained from a user study is very helpful. For example, it has been shown in Oviatt, DeAngeli, and Kuhn (1997) that users' written input precedes speech during a sequentially integrated multimodal command. They have also clarified the distribution of typical intermodal lags.

Semantic Fusion with Uncertain Inputs

The challenge of semantic fusion with uncertain inputs is to determine the unified meaning based on multimodal input fragments associated with probabilities. This is especially important for speech-centric MUI because the output of a speech recognizer is never certain. Note that the unification operation on the typed feature structure assumes that all input modalities are certain, and so they cannot be directly applied here. To fuse modalities with uncertainties, a *hybrid symbolic/statistical* architecture that combines statistical processing techniques with a symbolic unification-based approach is in need. This combined approach involves many factors when fusing the semantics. These factors include recognition accuracy of the individual modalities, the way of combining posterior probabilities, and the prior distribution of multimodal commands.

Note that a multimodal input gives rise to three different types of information overlay: nonoverlayed, overlayed and nonconflicting, and overlayed and conflicting. Nonoverlayed information indicates that the input (unimodal or multimodal) does not have any of the same information represented multiple times. This is the simplest condition. Overlayed and nonconflicting information refers to information segments that may have been represented multiple times without a conflict. The overlayed and conflicting information refers to the case that the information has been provided multiple times and conflicts. There are many approaches to resolving conflicting information in typed feature structure if no uncertainty is involved. The "unification" approach simply returns the value null when a conflict is detected. The "overlay" method returns the first argument when conflicting information is present. However, given that the semantic information from different modalities should not be equally trusted, a better conflicting information resolving approach can be found to handle input signals that may or may not be overlapped in their temporal delivery (Oviatt et al., 1997). Note that overlayed information may arise when inputs are from different modalities (e.g., speech and gesture), or when the same-type modality information occurs multiple times over an extended time frame. Both these two conditions need to be handled.

Conventionally, the probability of the merged feature structures is the cross product of the probabilities of individual feature structures based on the assumption that inputs are statistically independent with each other. In this section, we describe an alternative statistical approach that has been used in QuickSet (Wu, Oviatt, & Cohen, 1999). This approach uses the associative map to reduce the unification pairs and members-teams-committee (MTC) model to refine the multimodal integration process so that different weights are assigned to different modes and different constituents.

Associative map defines all semantically meaningful mapping relations that exist between different sets of constituents for each multimodal command. In its simplest form, it can be considered as a simple process of table lookup. For example, if an MUI consists of only the speech modality and the pen modality, we can build a two-dimensional table. If two inputs from different modalities can be fused, the value at the corresponding cell is 1; otherwise, the value is 0. The purpose of the associative map is to rule out considerations of those feature structures that cannot possibly be unified semantically.

Members-teams-committee weighs the contributions derived from different modality recognizers based on their empirically-derived relative reliabilities. MTC consists of multiple members, multiple teams, and a committee. "*members* are the individual recognizers that provide a diverse spectrum of recognition results (local posterior probabilities). Member recognizers can be on more than one team. Members report their results to their recognizer *team* leader, which then applies various weighting parameters to their reported scores. Furthermore, each team can apply a different weighting scheme, and can examine different subsets of data. Finally, the *committee* weights the results of the various teams, and reports the final recognition results. The parameters at each level of the hierarchy are trained from a labeled corpus." (Oviatt, et al., 2000, online version, p. 24).

CONCLUSION AND FUTURE DIRECTIONS

In this chapter, we discussed the importance of using the MUI in mobile devices, and described the state-of-the-art technologies in designing speech-centric MUI in mobile devices. Specifically, we discussed the noise robustness technologies, the reliable modality switching methods, the context-aware language model, and the robust language-understanding technologies that contribute to the usability of the speech modality. We also described the modality integration technologies that are important to improving the accuracy and throughput of the MUI. Although these technologies have greatly advanced the speech centric MUI design and development in the mobile devices, future research is needed in the following areas.

Microphone Array Processing

Noise robustness is still a challenging research area for speech-centric MUIs. Although many single-microphone noise robustness technologies (e.g., Deng, et al., 2002; Macho, et al. 2002) have been proposed to improve speech recognition accuracy under noisy environments, the progress so far is still limited. Given the continuous decrease in the hardware price, using microphone array on mobile devices is a trend to combat noisy acoustic conditions and to further decrease speech recognition errors. Microphone array algorithms, which take advantage of the received signal differences between microphones, can achieve noise suppression of 10-15 db effectively (Tashev & Malvar, 2005). Future research is needed for more efficient and effective algorithms using low-cost, low-quality microphone arrays that may be equipped in speech-centric mobile devices.

Error Handling Techniques

Fragile error handling continues to be a top interface problem for speech-centric MUI (Karat, Halverson, Horn, & Karat, 1999; Rhyne & Wolf, 1993; Roe & Wilpon, 1994). A great amount of research work needs to be done in developing graceful error-handling strategies in speech-centric MUI. First, new statistical methods need to be developed to reduce errors through mutual disambiguation between modalities. Second, new dialog strategies (e.g., mixed initiative) need to be developed to allow easy correction of the errors. Third, the system needs to be able to adapt to different environments and challenging contexts to reduce errors. Fourth, better robust speech recognition technologies need to be developed to increase the speech recognition accuracy under a wide range of environments.

Adaptive Multimodal Architectures

In most current MUI systems, their behaviors are predesigned by the developers. The system does not automatically learn to improve the performance as users use the system. Given that mobile devices are usually used by a single user, it is very important to develop adaptive MUI architectures.

For example, Oviatt (1999) showed that any given user's habitual integration pattern (simultaneous vs. sequential) is apparent at the beginning of their system interaction. When the user uses the system, the interaction pattern remains the same. An adaptive MUI system that can distinguish and utilize these patterns to improve the modality fusion could potentially achieve greater recognition accuracy and interactive speed. Another example is for the system to gradually change the behavior (e.g., automatically predict the user's next action) when the user changes from a novice to an experienced user.

Future research in this area would include what and when to adapt, as well as how (e.g., through reinforcement learning) to adapt MUI systems so that their robustness can be enhanced.

Mixed Initiative Multimodal Dialog

Most current speech-centric MUI systems are user initiative, where the user controls the dialog flow (for example, through push to talk). A user-initiative system can be modeled as a set of asynchronous event handlers. In a more advanced system, the system should also actively interact with the user to ask for missing information (which is called mixed initiative). For example, if the user wants to search for the phone number of a business using a mobile device and he/she forgets to mention the city and state information, the dialog system should automatically ask the user for that information through the multimodal output devices.

Future research should address the design and development of consistent and efficient conversational interaction strategies that can be used by different multimodal systems. Multimodal dialogue systems should be developed within a statistical framework (Horvitz, 1999) that permits probabilistic reasoning about the task, the context, and typical user intentions.

REFERENCES

Becker, N. (2001). *Multimodal interface for mobile clients*, Retrieved July 16, 2006, from http://citeseer.ifi.unizh.ch/563751.html

Bolt, R. A. (1980). Put-that-there: Voice and gesture at the graphics interface. *Computer Graphics, 14(3)*, 262-270.

Bregler, C., Manke, S., Hild, H., & Waibel, A. (1993). Improving connected letter recognition by lipreading. *Proceedings of the International Conference on Acoustics, Speech and Signal Processing, 1*, 557-560.

Cheyer, A., & Julia, L. (1995). Multimodal maps: An agent-based approach. *International Conference on Cooperative Multimodal Communication* (pp. 103-113).

Codella, C., Jalili, R., Koved, L., Lewis, J., Ling, D., Lipscomb, J., Rabenhorst, D., Wang, C., Norton, A., Sweeney, P., & Turk, C. (1992). Interactive simulation in a multi-person virtual world. *Proceedings of the Conference on Human Factors in Computing Systems* (pp. 329-334).

Cohen, P. R., Dalrymple, M., Moran, D. B., Pereira, F. C. N., Sullivan, J. W., Gargan, R. A., Schlossberg, J. L., & Tyler, S. W. (1989). Synergistic use of direct manipulation and natural language. *Proceedings of the Conference on Human Factors in Computing Systems* (pp. 227-234).

Cohen, P. R., Johnston, M., McGee, D., Oviatt, S., Pittman, J., Smith, I., Chen, L., & Clow, J. (1997). Quickset: Multimodal interaction for distributed applications. *Proceedings of the Fifth ACM International Multimedia Conference* (pp. 31-40).

Das, S., Bakis, R., Nadas, A., Nahamoo, D. & Picheny, M. (1993). Influence of background noise and microphone on the performance of the IBM TANGORA speech recognition system. *Proceedings of the IEEE International Conference on Acoustic Speech Signal Processing* (pp. 71-74).

Deligne, S., Dharanipragada, S., Gopinath, R., Maison, B., Olsen, P., & Printz, H. (2002). A robust high accuracy speech recognition system for mobile applications. *IEEE Transactions on Speech and Audio Processing, 10*(8), 551-561.

Deng, L., Wang, K., Acero, A., Hon, H., Droppo, J., Boulis, C., Wang, Y., Jacoby, D., Mahajan, M., Chelba, C., & Huang, X.D. (2002). Distributed speech processing in MiPad's multimodal user interface. *IEEE Transactions on Speech and Audio Processing*, *10*(8), 605-619.

Deng, L, & Yu, D. (2005). A speech-centric perspective for human-computer interface - A case study. *Journal of VLSI Signal Processing Systems (Special Issue on Multimedia Signal Processing)*, *41*(3), 255-269.

Dybkjaer, L., & Bernsen, N.O . (2001). Usability evaluation in spoken language dialogue system. *Proceedings of the Workshop on Evaluation for Language and Dialogue Systems, Association for Computational Linguistics 39th Annual Meeting and 10th Conference of the European Chapter*,,(pp. 9-18).

Fukumoto, M., Suenaga, Y., & Mase, K. (1994). Finger-pointer: Pointing interface by image processing. *Computer Graphics, 18(5),* 633-642.

Gong, Y. (1995). Speech recognition in noisy environments: A survey. *Speech Communication*, *16*, 261-291.

Gould, J. & Lewis, C. (1985). Design for usability: Key principles and what designers think. *Communications of the ACM*, *28*(3): 300-301.

Holzman, T. G. (1999). Computer-human interface solutions for emergency medical care. *Interactions, 6(3),* 13-24.

Hone, K. S., & Graham, R. (2001). Subjective assessment of speech system interface usability. *Proceedings of the Eurospeech Conference* (pp. 2083-2086).

Horvitz, E. (1999). Principles of mixed-initiative user interfaces. *Proceedings of the Conference on Human Factors in Computing Systems* (pp. 159-166).

Huang, X., Acero, A., Chelba, C., Deng, L., Droppo, J., Duchene, D., Goodman, J., Hon, H., Jacoby, D., Jiang, L., Loynd, R., Mahajan, M., Mau, P., Meredith, S., Mughal, S., Neto, S., Plumpe, M., Stery, K., Venolia, G., Wang, K., & Wang. Y. (2001). MIPAD: A multimodal interaction prototype. *Proceedings of the International Conference on Acoustics*, *Speech*, *and Signal Processing*, *1*, 9-12.

Huang, X., Acero, A., Chelba, C., Deng, L., Duchene, D., Goodman, J., Hon, H., Jacoby, D., Jiang, L., Loynd, R., Mahajan, M., Mau, P., Meredith, S., Mughal, S., Neto, S., Plumpe, M., Wang, K., & Wang, Y. (2000). MIPAD: A next generation PDA prototype. *Proceedings of the International Conference on Spoken Language Processing, 3*, 33-36.

Hsu B.-J., Mahajan M., & Acero A. (2005). *Multimodal text entry on mobile devices.* The ninth bi-annual IEEE workshop on Automatic Speech Recognition and Understanding (Demo). Retrieved July 20, 2006, from

http://research.microsoft.com/~milindm/2005-milindm-ASRU-Demo.pdf

Johnston, M. (1998). Unification-based multimodal parsing. *Proceedings of the International Joint Conference of the Association for Computational Linguistics and the International Committee on Computational Linguistics* (pp. 624-630).

Karat, C.-M., Halverson, C., Horn, D., & Karat, J. (1999). Patterns of entry and correction in large vocabulary continuous speech recognition systems. *Proceedings of the International Conference for Computer-Human Interaction* (pp. 568-575).

Kay, M. (1979). Functional grammar. *Proceedings of the Fifth Annual Meeting of the Berkeley Linguistics Society* (pp. 142-158).

Kumar, S., & Cohen, P. R. (2000). Towards a fault-tolerant multi-agent system architecture. *Fourth International Conference on Autonomous Agents* (pp. 459-466).

Li, X., Malkin J., & Bilmes, J. (2006). A high-speed, low-resource ASR back-end based on custom arithmetic. *IEEE Transactions on Audio, Speech and Language Processing, 14*(5), 1683-1693.

Lockwood, P., & Boudy, J. (1992). Experiments with a nonlinear spectral subtractor (NSS), hidden Markov models and the projection for robust speech recognition in cars. *Speech Communication, 11*(2-3), 215-28.

Macho, D., Mauuary, L., Noé, B., Cheng, Y. M., Ealey, D., Jouvet, D., Kelleher, H., Pearce, D., & Saadoun, F. (2002). Evaluation of a noise-robust DSR front-end on Aurora databases. *Proceedings of the International Conference on Spoken Language Processing* (pp. 17-20).

Minsky, M. (1975). A framework for representing knowledge. In P. Winston (Ed.), *The psychology of computer vision* (pp 211-277). New York: McGraw-Hill.

Neal, J. G., & Shapiro, S. C. (1991). Intelligent multimedia interface technology. In J. Sullivan & S. Tyler (Eds.), *Intelligent user interfaces* (pp.11-43). New York: ACM Press.

Nielsen, J. (1993). *Usability engineering.* San Diego: Academic Press .

Norman, D. A. & Draper, S. W. (Eds.). (1986). *User-centered system design: New perspectives on human-computer interaction*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Oviatt, S. L. (1999). Ten myths of multimodal interaction, *Communications of the ACM, 42* (11), 74-81.

Oviatt, S. L., Bernard, J., & Levow, G. (1999). Linguistic adaptation during error resolution with spoken and multimodal systems. *Language and Speech* (special issue on *Prosody and Conversation), 41*(3-4), 415-438.

Oviatt, S. L., Cohen, P. R., Wu, L., Vergo, J., Duncan, L., Suhm, B., Bers, J., Holzman, T., Winograd, T., Landay, J., Larson, J., & Ferro, D. (2000). Designing the user interface for multimodal speech and gesture applications: State-of-the-art systems and research directions. *Human Computer Interaction*, 263-322. (online version), Retrieved July 20, 2006, from http://www.cse.ogi.edu/CHCC/Publications/designing_user_interface_multimodal_speech_oviatt.pdf

Oviatt, S. L., DeAngeli, A., & Kuhn, K. (1997). Integration and synchronization of input modes during multimodal human-computer interaction. *Proceedings of Conference on Human Factors in Computing Systems (CHI'97)* (pp. 415-422).

Oviatt, S. L. & Olsen, E. (1994). Integration themes in multimodal human-computer interaction. In Shirai, Furui, & Kakehi (Eds.), *Proceedings of the International Conference on Spoken Language Processing, 2*, 551-554.

Oviatt, S. L., & vanGent, R. (1996). Error resolution during multimodal human-computer interaction. *Proceedings of the International Conference on Spoken Language Processing, 2*, 204-207.

Pavlovic, V., Berry, G., & Huang, T. S. (1997). Integration of audio/visual information for use in human-computer intelligent interaction. *Proceedings of IEEE International Conference on Image Processing* (pp. 121-124).

Pavlovic, V., & Huang, T. S., (1998). Multimodal prediction and classification on audio-visual features. *AAAI'98 Workshop on Representations for Multi-modal Human-Computer Interaction*, 55-59.

Ravden, S. J., & Johnson, G. I. (1989). *Evaluating usability of human-computer interfaces: A practical method.* Chichester: Ellis Horwood.

Reeves, L. M., Lai, J., Larson, J. A., Oviatt, S., Balaji, T. S., Buisine, S., Collings, P., Cohen, P., Kraal, B., Martin, J. C., McTear, M., Raman, T. V., Stanney, K. M., Su, H., & Wang, Q. Y. (2004). Guidelines for multimodal user interface design. *Communications of the ACM – Special Issue on Multimodal Interfaces, 47*(1), 57-59.

Rhyne, J. R., & Wolf, C. G. (1993). Recognition-based user interfaces. In H. R. Hartson & D. Hix (Eds.), *Advances in Human-Computer Interaction, 4*, 191-250.

Roe, D. B., & Wilpon, J. G. (Eds.). (1994). *Voice communication between humans and machines.* Washington, D.C: National Academy Press.

Rubin, P., Vatikiotis-Bateson, E., & Benoit, C. (1998). Special issue on audio-visual speech processing. *Speech Communication*, *26* (1-2).

Schwartz, D. G. (1993). *Cooperating heterogeneous systems: A blackboard-based meta approach*. Unpublished Ph. D. thesis, Case Western Reserve University.

Shaikh, A., Juth, S., Medl, A., Marsic, I., Kulikowski, C., & Flanagan, J. (1997). An architecture for multimodal information fusion. *Proceedings of the Workshop on Perceptual User Interfaces,* 91-93.

Stork, D. G., & Hennecke, M. E. (Eds.) (1995). *Speechreading by humans and machines.* New York: Springer Verlag.

Suhm, B. (1998). *Multimodal interactive error recovery for non-conversational speech user interfaces.* Ph.D. thesis, Fredericiana University.

Tashev, I., & Malvar, H. S. (2005). A new beamformer design algorithm for microphone arrays. *Proceedings of International Conference of Acoustic, Speech and Signal Processing*, *3*, 101-104.

Varga, I., Aalburg, S., Andrassy, B., Astrov, S., Bauer, J.G., Beaugeant, C., Geissler, C., & Hoge, H. (2002). ASR in mobile phones - an industrial approach, *IEEE Transactions on Speech and Audio Processing*, *10*(8), 562- 569.

Vo, M. T., Houghton, R., Yang, J., Bub, U., Meier, U., Waibel, A., & Duchnowski, P. (1995). Multimodal learning interfaces. *Proceedings of the DARPA Spoken Language Technology Workshop.* Retrieved July 20, 2006 from

http://www.cs.cmu.edu/afs/cs.cmu.edu/user/tue/www/papers/slt95/paper.html

Vo, M. T., & Wood, C. (1996). Building an application framework for speech and pen input integration in multimodal learning interfaces. *Proceedings of IEEE International Conference of Acoustic, Speech and Signal Processing, 6,* 3545-3548.

Wang, J. (1995). Integration of eye-gaze, voice and manual response in multimodal user interfaces. *Proceedings of IEEE International Conference on Systems, Man and Cybernetics* (pp. 3938-3942).

Wang, K. (2004). A detection based approach to robust speech understanding. *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing, 1,* 413-416.

Wang, Y. (2001). Robust language understanding in MiPAD. *Proceedings of the Eurospeech Conference* (pp. 1555-1558).

Wang, Y., Acero, A., & Chelba, C. (2003). Is word error rate a good indicator for spoken language understanding accuracy? *Proceedings of the Workshop on Automatic Speech Recognition Workshop and Understanding* (pp 577-582).

Wu, L., Oviatt, S., &. Cohen, P. (1999). Multimodal integration-A statistical view. *IEEE Transactions on Multimedia*, *1*(4), 334-341.

Yu, D., Ju, Y. C., Wang, Y., & Acero, A. (2006). N-gram based filler model for robust grammar authoring. *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing1,* 565-569.

Yu, D., Wang, K., Mahajan, M., Mau, P., & Acero, A. (2003). Improved name recognition with user modeling. *Proceedings of Eurospeech* (pp. 1229-1232).

KEY TERMS

Modality: A communication channel between human and computer, such as vision, speech, keyboard, pen, and touch.

Modality Fusion: A process of combining information from different input modalities in a principled way. Typical fusion approaches include early fusion, in which signals are integrated at the feature level, and late fusion, in which information is integrated at the semantic level.

Multimodal User Interface: A user interface with which users can choose to interact with a system through one of the supported modalities, or multiple modalities simultaneously, based on the usage environment or preference. Multimodal user interface can increase the usability because the strength of one modality often compensates for the weaknesses of another.

Push to Talk: A method of modality switching where a momentary button is used to activate and deactivate the speech recognition engine.

Speech-Centric Multimodal User Interface: A multimodal user interface where speech is the central and primary interaction modality.

Typed feature Structure: An extended, recursive version of attribute-value type data structures, where a value can, in turn, be a feature structure. It indicates the kind of entity it represents with a type, and the values with an associated collection of feature-value or attribute-value pairs. In the typed feature structure, a value may be nil, a variable, an atom, or another typed feature structure.

User-Centered Design: A design philosophy and process in which great attention is given to the needs, expectations, and limitations of the end user of a human-computer interface at each stage of the design process. In the user-centered design process, designers not only analyze and foresee how users are likely to use an interface, but also test their assumptions with actual users under real usage scenario.

Chapter XXIX
Model-Based Target Sonification in Small Screen Devices:
Perception and Action

Parisa Eslambolchilar
University of Wales, UK

Andrew Crossan
University of Glasgow, UK

Roderick Murray-Smith
University of Glasgow, UK
Hamilton Institute, NUI Maynooth, Ireland

Sara Dalzel-Job
University of Glasgow, UK

Frank Pollick
University of Glasgow, UK

ABSTRACT

In this work, we investigate the use of audio and haptic feedback to augment the display of a mobile device controlled by tilt input. The questions we answer in this work are: How do people begin searching in unfamiliar spaces? What patterns do users follow and which techniques are employed to accomplish the experimental task? What effect does a prediction of the future state in the audio space, based on a model of the human operator, have on subjects' behaviour? In the pilot study we studied subjects' navigation in a state space with seven randomly placed audio sources, displayed via audio and vibrotactile modalities. In the main study, we compared only the efficiency of different forms of audio feedback. We ran these experiments on a Pocket PC instrumented with an accelerometer and a headset. The accuracy *of selecting, exploration density, and orientation of each target was measured. The results quantified the changes brought by predictive or "quickened" sonified displays in mobile, gestural interaction. Also, they highlighted subjects' search patterns and the effect of a combination of independent variables and each individual variable in the navigation patterns.*

INTRODUCTION

One of the main goals of interaction design is to make the interfaces as intuitive as possible. In our everyday environments, humans receive a variety of stimuli playing upon all senses, including aural, tactile, and visual, and we respond to these stimuli. Even though hearing and vision are our two primary senses, most of today's interfaces are mainly visual.

Visual interfaces have crucial limitations in small-screen devices. These devices have a limited amount of screen space on which to display information. Designing interfaces for mobile computers/phones is problematic, as there is a very limited amount of screen resource on which to display information, and users' need to focus on the environment rather than the interface (so that they can look where they are going) so output is limited (Blattner, Papp, & Glinert, 1992; Brewster, 1997; Brewster & Murray, 1998; Johnson, Brewster, Leplatre, & Crease, 1998; Kramer, Walker, Bonebright, Cook, Flowers, Miner, 1999; Rinott, 2004; Smith & Walker, 2005; Walker & Lindsay, 2006); also, low graphics resolution and further constrain the freedom of interface designers. In new generations of mobile phones (e.g., iPhone) with high graphics resolution, power consumption for graphics rendering is high, which can adversely affect battery life; also, large screens can lead to physical robustness issues, as well as being very demanding of user attention in mobile scenarios.

One way around these problems would be sonically enhanced interfaces that require less or no visual attention and therefore, the size of the visual display and the portable device can be decreased; also, auditory interfaces potentially interfere less in the main activity in which the user is engaged. Consequently, the user may be able to perform more than one task at a time, such as driving a car while using a telephone or grabbing a cup of coffee while waiting for a mobile phone to finish downloading an image. Auditory feedback can often be a necessary complement, but also a useful alternative to visual feedback. When designing a mobile electronic device, it is difficult to predict all possible scenarios when it might be used. Obviously, visual feedback is preferred in many situations such as in noisy environments or when the user has to concentrate on a listening task. However, as there might be numerous occasions when a user cannot look at a display, versatile devices such as mobile phones or handheld computers benefit from having flexible interfaces.

Novel Interaction and Continuous Control

In the past 10 years many researchers have focused on tilt-based inputs, and audio and haptic outputs in mobile HCIs (Dong, Watters, & Duffy, 2005; Fallman, 2002a, 2002b; Harrison & Fishkin, 1998; Hinckley, Pierce, Horvitz, & Sinclair, 2005; Oakley, Ängeslevä, Hughes, & O'Modhrain, 2004; Partridge, Chatterjee, Sazawal, Borriello, & Want, 2002; Rekimoto, 1996; Sazawal, Want, & Borriello, 2002; Wigdor & Balakrishnan, 2003). The results of these researches have proved one-handed control of a small screen device needs less visual attention than two-handed control and multimodality in the interaction can compensate for the lack of screen space. So these novel interaction techniques, that is, gesture recognition, and audio and haptic devices, are characterised by the significance of the temporal aspect of interaction and in such an emerging environment, the interaction is no longer based on a series of discrete steps, but on a continuous input/output exchange of information that occurs over a period of time at a relatively high rate, somewhat akin to vision based or audio/haptic interfaces, which we may not model appropriately as a series of discrete events (Doherty & Massink, 1999; Faconti & Massink, 2001).

Novel interaction techniques with computers and handheld devices are examples of interactive dynamic systems, and development of these systems explores a range of possible solutions for overcoming some problems of development on computing devices, including the limited source of input/output devices, adaptability, predictability, disturbances, and individual differences. We explicitly include dynamics because we experience our environment in the way we want it by our actions or behaviour. Thus, we control what we perceive and while, in principle, interaction with handheld devices is rich in the variety of tasks supported, from computation and information storage to sensing and communication, we are dependent on the display of feedback (either visual, audio, or haptic) to help us pursue our sometime constantly changing goals feedback, which may influence a user's actions as more information becomes available (Doherty & Massink, 1999; Faconti & Massink, 2001). So developing interaction for such devices is closely related to the engineering of mobile interfaces based on dynamics.

Control Theory and Fitts' Law

A branch of control theory that is used to analyse human and system behaviour when operating in a tightly coupled loop is called *manual control theory* (Jagacinski & Flach, 2003; Poulton, 1974). The theory is applicable to a wide range of tasks involving vigilance, tracking and stability, and so forth. The general approach followed in manual control theory is to express the dynamics of combined human and controlled element behaviour as a set of linear differential equations in the time domain (Poulton, 1974). Several models include human-related aspects of information processing explicitly in the model, such as delays for visual process, motor-nerve latency, and neuromotor dynamics (Jagacinski & Flach, 2003). Control theory can be linked to Fitts' Law (Fitts, 1954; MacKenzie & Ware, 1993; Mackinlay, Robertson, & Card, 1991) by viewing the pointing movements towards the target as a feedback control loop based on visual input, and the limb as a control element (Bootsma, Fernandez, & Mottet, 2004; Crossman & Goodeve, 1983; Hoffmann, 1991; Jagacinski & Flach, 2003; Langolf, Chaffin, & Foulke, 1976).

This research outlines the use of model-based sonification to shape human action when users interact with small devices based on auditory feedback. In this work, we investigate the usability of nonspeech sounds and haptic feedback to augment the display of a mobile device controlled by a gesture input. Nonspeech sound has advantages over speech in that it is faster as well as language independent. We look at control strategies of users in browsing the audio/haptic state space. We also suggest one possible way of improving performance based on models of human control behaviour in a few example applications.

BACKGROUND

The single audio output channel has been little used to improve interaction in mobile devices. Speech sounds are, of course, used in mobile phones when calls are being made, but are not used by the telephone to aid the interaction with the device (Blattner et al., 1992; Brewster, 2002; Gaver, Smith, & O'Shea, 1991; Smith & Walker, 2005; Walker & Lindsay, 2006). Nonspeech sounds and vibrotactile devices are used for ringing tones or alarms but again, do not help the user interact with the system beyond this. Some signals provide feedback that some event has been successful, such as when buttons are pressed or devices are switched on. Selecting items with a stylus in PDAs without tactile feedback is often confusing for users because it is hard to know whether they have hit the target or not, especially if used in a mobile setting (Brewster, 2002). In this case, vibrators in mobile phones could be a good haptic feedback. It assures the user that s/he is in the target, and if the user wants to select a target, s/he can then press a key in the vibration area to select it.

If using continuous sounds as opposed to the more common brief signals, auditory interfaces do not need to be more transitory than visual interfaces. However, such sounds probably benefit from being quite discreet. While, most existing sound feedback today occurs in the foreground of the interface, subtle background sounds can be a useful complement in advanced auditory interfaces. Films and computer games generally make use of music and sound effects. Film sound theorist Michel Chion (Chion, 1994) has made the following statement concerning sound in film: *there is no soundtrack*. An extreme statement coming from a researcher of sound, Chion means that there is no way to separate the auditory and visual channels of a film. We experience them only through a unified sense, which he terms "audio-vision."

In a similar way, an interface that uses sound cleverly can enhance the user's immersion and improve interaction. Gaver (1997) found that during an experimental process control task, the participants' engagement increased when provided with relevant sound feedback. There is now evidence that sound can improve interaction and may be very powerful in small screen devices (Brewster, 2002). If the possibility of conveying information sonically were used to its full potential, it would be a powerful complement to visual interfaces (Brewster & Murray, 1998). A strong argument against the use of sound in interfaces is that it easily can become annoying, both for the user and other people around them, since it is more intrusive than visual impressions. It is not useful in noisy environments, for instance, train stations, undergrounds, and so forth. However, by skilfully designing auditory interfaces or using haptic feedback, this can be avoided, and interaction with machines can become easier and hopefully more pleasant.

The most advanced auditory/haptic feedback seems to exist in computer games and multimedia products. Gaver (1997) claims that memory limitations in the technical product are one reason why sound feedback has not been used on a larger scale. Until quite recently, it has been too expensive computationally to use sound of good quality in computers and handheld devices. Today, only lightweight electronic devices, such as mobile phones or handheld computers, have limited memory capacities, although this is rapidly changing with the development of memory cards and effective compression algorithms for sound. However, nowadays these devices give various choices of discrete audio/haptic ring tones and alarms and to their users. The potential to use sound and haptic in small electronics is growing fast.

MODEL-BASED SONIFICATION

As there are many ways in which sound can be employed in interfaces, it is important to define the purposes of every sound at an early stage in the design process. A sound that conveys crucial information should have different attributes to one that serves as a complement to visual information. It is important to distinguish between two very different approaches (Chion, 1994): the practical and the naturalistic approach. The "practical" approach to auditory interfaces deals with sound as the main feedback. This can be the case when designing interfaces for visually impaired people, who must rely on sound feedback to provide sufficient assistance in performing a task. Furthermore, sound is often the only means of communication when using a portable hands-free device with a mobile phone. Auditory interfaces based on a practical approach should be comprehensive and simple (Brewster, 2002; Brewster & Murray, 1998; Smith & Walker, 2005; Walker & Lindsay, 2006). The drawback of this approach is sound might be noisy and tiresome over time. The "naturalistic" view regards sound mainly as a complement to a visual interface. A naturalistic interface combines sound and vision in a way as similar as possible to corresponding phenomena in the natural world. Such auditory interfaces are supposed to enhance interaction between the user and a machine, especially in situations where the visual interface is ineffective on its own. Sounds that complement a visual interface can generally be subtle background events that do not disturb. In a way, such sounds correspond to the background music of films, since they convey information to the audience without interfering with the main events. Sound feedback based on the naturalistic strategy is thus very subtle, and might only be recognised subconsciously. The focus of this work is on the "practical" approach.

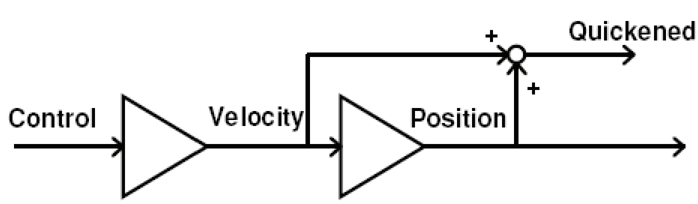
Sonification is a method suggested in "practical" domain, which is defined as the use of nonspeech audio to convey information. More specifically, sonification is the transformation of data relations into perceived relations in an acoustic signal for the purposes of facilitating communication or interpretation (Gaver, 1989). Many of the major current research areas in sonification are similar in that they focus on the identification of applications for which audition provides advantages over other modalities, especially for situations where temporal features are important or the visual modality is overtaxed. The main issues that will move sonification research forward include (1) mapping data onto appropriate sound features like volume, pitch, timbre, (2) understanding dynamic sound perception, (3) investigating auditory streaming, (4) defining and categorising salience in general auditory contexts, and understanding where highly salient sonic events or patterns can surpass visual representations in data mining, and (5) developing multimodal applications of sonification (Kramer et al., 1999); sonification is a way to help in the exploration of complex data. Various kinds of information can be presented using sonification, simply by using different acoustic elements. This information has been organised in Hermann, Hansen, and Ritter (2000).

Studies such as Cook et al. (2002) and Cook and Lakatos (2003) have investigated the human ability to perceive various physical attributes of sound sources, and have proved that feature-based synthesis is of use in studying the low-level acoustical properties that human listeners use to deduce the more complex physical attributes of a sound's source. The generated sounds from a set of features are correlated with the listener's perception of, for example, size, speed, or shape of the source. Two methods of sonification have been used in this chapter, the Doppler effect and derivative volume adaptation. Both of these methods create a continuous sound for each data point. Thus, the relative position to the targets is perceived by a change of volume when passing the data point and pitch shift for Doppler effect as well. From the data points obtained in this way, we may be able to discover consistent relationships between acoustical and human-generated features that can be used to predict how a sound manifesting certain acoustic feature values will be perceived.

Quickening

"Quickening" is a method for reducing the difficulty of controlling second-order or higher-order systems, by changing the display to include predictions of future states, that was proposed by Birmingham and Taylor (1954), and is reviewed in Jagacinski and Flach (2003). A quickened display for an acceleration control system like the system described in this chapter shows the user a weighted combination of position and velocity (see Figure 1). This weighted summation effectively anticipates the future position of the system. It can greatly improve human performance in controlling these

Figure 1. A block diagram for a second-order system with a quickened display. The output to the quickened display is the sum of position and velocity. Effectively, the quickened display projects the output into the future based on the current velocity.



systems. Quickening in general is a prediction of the future state of the system based on the current state vector (for example position, velocity, acceleration) and a model of system behaviour and expected user action.

An example of this is based on the Doppler effect, which highlights the user's approach to a target, or a target's movement from the current state. Another example could be derivative of volume of sound source. When the user is further from the audio source, the sound is quieter than when the user is close to it. Another predictive method that has been investigated in Williamson, Strachan, and Murray-Smith (2006) is *Monte Carlo* simulation in a tilt-controlled navigation system.

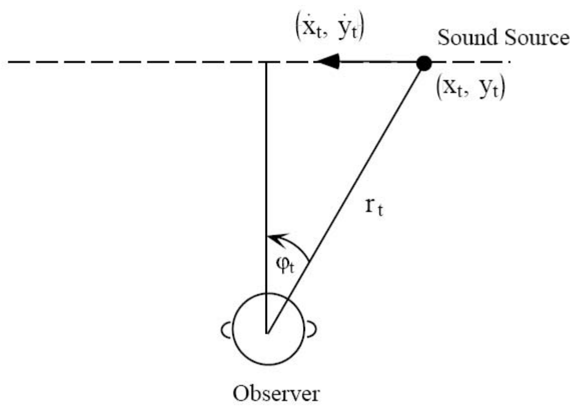
Doppler Effect

The auditory system is responsible for constructing a map of the auditory scene around us, using information from audio input, that is, sound localisation (Bregman, 1990; Smith, 2004). There are various types of cues that humans can use to localise the position of a sound source. These cues can be divided into monaural and binaural cues. The two different types of monaural cue are loudness and Doppler shift. The loudness cue relies on the fact that when a sound source is far away, it is quieter than when it is close by. The Doppler shift corresponds to a frequency shift associated with a sound source moving through a homogeneous medium (Smith, 2004). Pressure wave crests emerge from the sound source at intervals corresponding to the acoustic wavelength. Each crest spreads spherically out from the point of origin at the speed of sound *c* (Figure 2). The successively generated spheres of wave crests are closer together ahead of the sound source but farther apart behind the source. For a stationary observer, the measured frequency corresponds to the number of crests per unit time, so the composite frequencies will be higher when the observer is in front of the moving sound source, and less when behind the moving sound source (Hermann et al., 2000; Hermann & Ritter, 1999). A familiar example is the shift in frequency of an ambulance siren as the vehicle approaches, passes, and then recedes. The well-known lawful dependence of the Doppler shifted frequency, here denoted Ψ_t, on velocity of the sound source relative to an observer is:

$$\Psi_t = f(1 + \frac{v}{c}\cos\Phi_t) \qquad (1)$$

where *f* is the intrinsic frequency of the sound source, *v* is the velocity magnitude (speed), and *c* is the speed of sound. The shifted frequency Ψ_t depends only on the velocity component directed toward the observer with angle Φ_t (see Figure 2). The shifted frequency has the maximum value when Φ_t is zero. As this angle reaches 90°, all motion is across the line of hearing and the Doppler shift is zero. This result holds true regardless of

Figure 2. The geometry for the Doppler shift of a moving sound source relative to an observer.



the time history of the trajectory (Jenison, 1997). These aural cues can be used to navigate through the virtual environment on a Pocket PC.

In the next sections we present pros and cons of different quickened methods and control strategies in browsing the state-space on a mobile device using tilt-input.

EXPERIMENT

Goals

There is a concept of accuracy explored in this work. The type of accuracy that is under primary consideration in this study is the capability of subjects to accurately identify audio sources in large audio data sets with a PDA and tilt-sensor using sound only. In navigating a computer display of data visually, accuracy is seldom a concern. Using a scrollbar or clicking a 10×10 pixel icon using one's vision is trivial from the perspective of the accuracy needed to accomplish this task (Holmes, 2005). Designers of auditory displays, on the other hand, are in need of research into the accuracy that is possible in this environment. Establishing the accuracy with which humans can navigate using sound alone is an early step in integrating sound into a multi-modal information system.

The other questions we answer in this work are: How do people begin searching in unfamiliar spaces? What patterns or techniques are employed to accomplish the experimental task? How will predicting the future state in the audio space change subjects' accuracy in targeting?

Apparatus

The experiment was conducted on a Pocket PC (hp5450), running windows CE, with a 240×320 resolution, colour display, an accelerometer Xsens P3C, 3 degree-of-freedom, attached to the serial port, which allows the users to navigate through the environment by tilting the device, and a stereo headset (Figure 3). The built-in vibrator unit in the Pocket PCs provides the haptic feedback in the experiment.

The experiment was written using the *FMOD* API (version 3.70CE)(FMOD, 2004), a visual programming environment with an object-oriented language (Embedded Visual C++) used primarily to manipulate and control sound production and *GapiDraw* (version 2.04) (GAPIDraw, 2004), a runtime add-in to *FMOD* used to generate real-time Pocket PC graphics. *FMOD,* and *GapiDraw* are available for free under the condition of the GNU public license (GPL).

Using *FMOD* and *GAPI*, an interface was developed with the following parameterisations: speed of sound, 340ms $^{-1}$, Doppler factor, 1.0, distance scale 100.0, minimum audible distance 80m, full volume (255)(minimum volume is 0

Figure 3. Left- Pocket PC, Accelerometer and experiment I running on the system (target sound sources displayed, for illustrative purposes). Right- A user interacting with the system.

and max volume is 255 in *FMOD*), and maximum audible distance 8000m. Each pixel on the display represents 100 metres. An empty window (240 × 320 pixels) was centred on the screen. Audio sources represented by small (10 × 10) speaker icons are shown on the screen only for training (Figure 3). In the main experiment, sound sources are hidden and an empty window is shown on the screen. Only the cursor, represented by a small (10 × 10) ear icon, is visible in both training and main experiment.

Experiment I

We first conducted a pilot study with 12 subjects, 3 women and 9 men, all sighted, with a mean age of 29 years. Four participants were research fellows, and the rest were postgraduate students at the NUIM campus. All but one of the participants had neither experience of using Pocket PCs nor with accelerometer-based interfaces. Two of them were left-handed (Eslambolchilar, Crossan, & Murray-Smith, 2004).

Task and Stimuli

The task in this study was to select the centre of individual targets that appear (in audio but not visually) in different locations on the screen as accurate as possible. The individual targets are audible when the cursor is in their locality, and they have full volume only in the centre of the target (imagine a Gaussian distribution of the volume centred on the target). For each target, a vibration feedback has been assigned and whenever the user is in very close distance to the target, 10 pixels, s/he feels the vibration continuously. Our aim in using the vibration in this task is the vibration assures the user that s/he is very close to the centre of the target.

First, participants were asked to sit on a chair in a quiet office and were equipped with a headset and a Pocket PC in their palm. Then they were informed about the functioning of the accelerometer, Doppler effect, and the procedures of the experiment, in order to reduce the chance of any terminological misunderstanding. Subjects were asked to move the cursor to audio targets by tilting the PDA, and to select them by pressing a key on a small keyboard of the PDA. They were told to emphasise accuracy over speed.

Design

There were four experimental conditions: (1) No Doppler effect, no vibration feedback (2) No Doppler effect, vibration feedback, (3) Doppler effect, no vibration feedback, and (4) Doppler effect, vibration feedback. The participants performed the conditions in a counterbalanced order. This resulted in 12 different orders of experiments for participants. In each experiment, seven audio sources were used (a selection of different music) summarised in Table 1.

Visualisation

Matlab was used for visualising the logged experimental data. We use a number of techniques for investigating the users' behaviour in these experiments.

Audio and exploration Density Plots

These plots show the audio density (in pixels) at different points in the 2-D space (Figure 4 (Left)). The contour indicates the density of the sum of the amplitude of the mixture components associated with the different audio tracks. The exploration

Table 1. Audio sources in the first experiment in all conditions

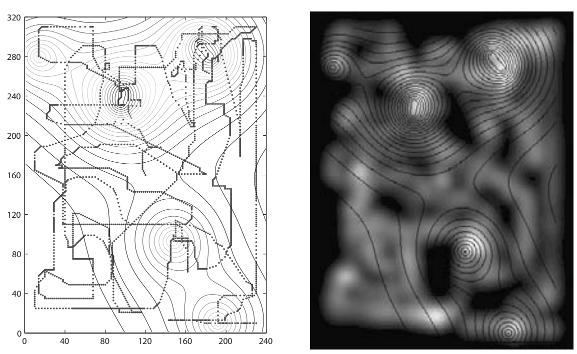
Target Index	Music Type
1	Hip-Hop
2	Celtic
3	Arabic
4	Country
5	Jazz
6	Farsi
7	Opera

density plot for visualisation of cursor trajectories, used previously in Williamson and Murray-Smith (2004), has been used here, which plots a density around the trajectory, which is a function of the position and the length of time spent in that position. These plots give some indication of how users navigated when completing the task. An example is given in Figure 4(Right). This plot is created by placing a Gaussian distribution centred on the (x,y) position of the cursor for each point in the log file, with standard deviation proportional to that used in the audio sources. The Gaussians are summed for each pixel, and the resulting image gives an impression of the areas of the input space that were explored, and how long the user spent in them. The image can be summarised numerically by counting the percentage of pixels greater than a selected threshold *e*. In this experiment *e*=5.0. The image's resolution is 240 by 320 pixels (Eslambolchilar et al., 2004).

Distance to the Target

Whenever the user feels s/he is at the target, s/he presses a key indicating the selection of the target. For each selection made by the user, the distance to the nearest target is calculated as below, and recorded.

Figure 4. Left - The cursor trace of the 4th participant in the "no Doppler-no vibration" condition, is plotted over the density of the local audio amplitude of the different tracks. Right - the density contour plot and cursor trajectory density indicating the exploration of the space by the same participant in the same condition.



$$Dist = \sqrt{(x_{source} - x_{selected})^2 + (y_{source} - y_{selected})^2} \tag{2}$$

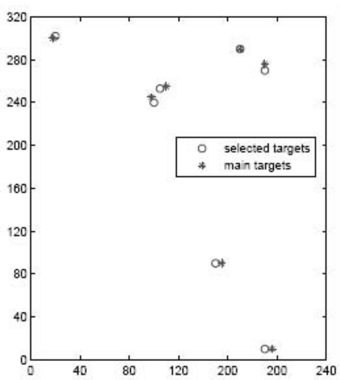
An example plot is shown in Figure 5. The distance to the location of the target (in pixels) gives some insight into the acuity with which the location can be perceived with the given display.

Results

Search Patterns Observed

In looking at the audio and exploration density plots, we are not attempting to establish a link between the search pattern used and the resulting measurement of accuracy. We simply make a subjective classification and qualitative assessment of the types of search patterns employed to accomplish the task. A subject may employ one of the search techniques and still not be very accurate, or they may be very accurate in spite of using no detectable systematic pattern. However, this factor gives an indication about the ease with which the audio environment could be clearly perceived by participants. In a clear and easy to navigate environment, with appropriate feedback, this should be similar to the density of targets, and linked to the smoothing used.

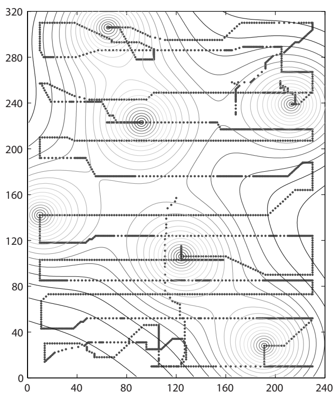
Figure 5. Hidden target positions (circles), and points selected by user 4 in the "no Doppler-no vibration" condition, as the best guess (crosses)



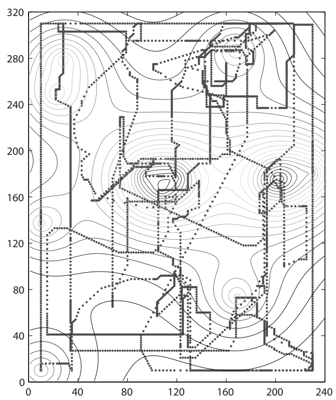
Some of the terms and their basic definitions used here are taken from search theory, a subfield within operations research (Patrol, 1999). The patterns developed by search theory are visual search patterns of physical space, but there is some crossover in the types of patterns used in the auditory interface used in the experiment to search in a virtual space.

1. **Parallel sweep:** The parallel sweep is used when uniform coverage of an area is desired and the area is unfamiliar. It is an efficient method of searching a large area in a minimum amount of time. Several subjects used the horizontal parallel sweep, "raster scan," similar to the one seen in Figures 6(a) or 7. This pattern can be related to the text-reading pattern we learn in the childhood.
2. **Quadrant search:** The quadrant search pattern is one in which the searcher mentally breaks down the screen into quadrants to divide the area into a more manageable size. Within the quadrants, the searcher may use another pattern to search each quadrant, such as a parallel sweep (Figure 8(a)).
3. **Sector search:** A sector search pattern begins once the approximate location of the target is located. In this pattern, the searcher explores out from the approximate location of the target and returns again, then conducts another exploration in another area, and returns again. This is repeated until they

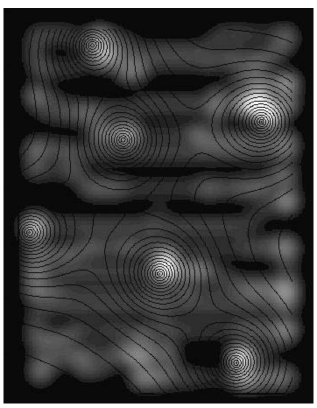
Figure 6. (Left) The traces of the cursor for participant 12 in "no Doppler with vibration" experiment (a) and its exploration density plot (c), (Right) The traces of the cursor for participant 6 in "Doppler with vibration" experiment (b) and exploration density plot of this experiment (d)



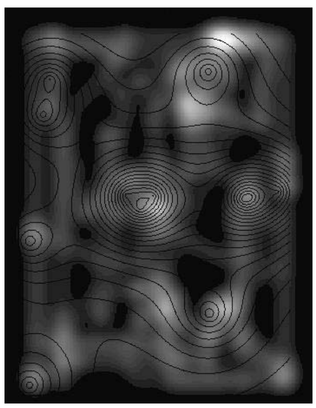
(a)



(b)



(c)



(d)

are confident that the space is adequately explored (Figure 8(b)).

4. **Perimeter search:** The perimeter search is one in which the boundaries of the space are explored, but little or none of the middle is traversed. The pattern of search can be a circle to circumscribe the border or a square-shaped pattern turning at a 90° angle. This type of search pattern would typically lead to inaccuracy given that none of the targets are located at the perimeter. This search pattern was not observed in this research.
5. **No formulaic search:** For some searchers, no discernable systematic technique was employed in exploring the space to accomplish the task. For these search patterns, there is no attempt to thoroughly explore the information space. Figure 8(c) illustrates the path used in the only trial to actually select the target exactly.

The search patterns of each subject were analysed to see if there were any tendencies based on demographic characteristics; 48 total patterns were analysed. The most common technique employed

Figure 7. (Left) The traces of the cursor for participant 5 in "Doppler with vibration" experiment, (Right) The traces of the cursor for participant 9 in "no Doppler with vibration" experiment.

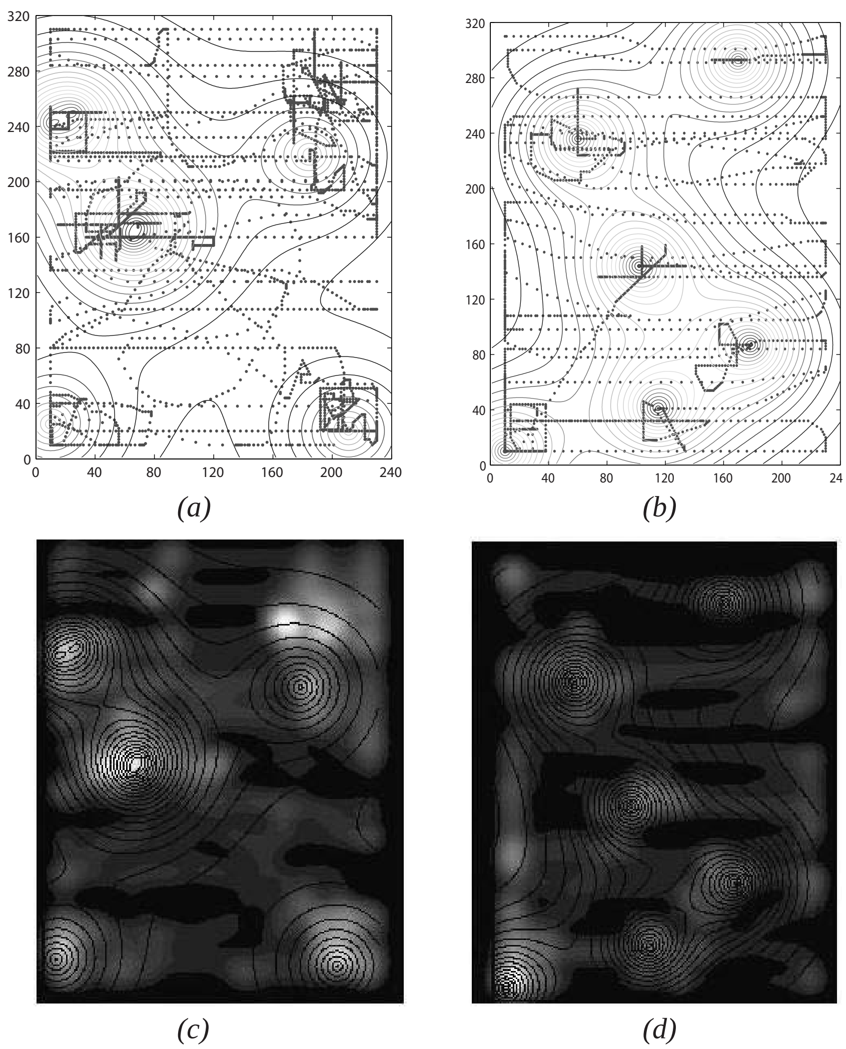
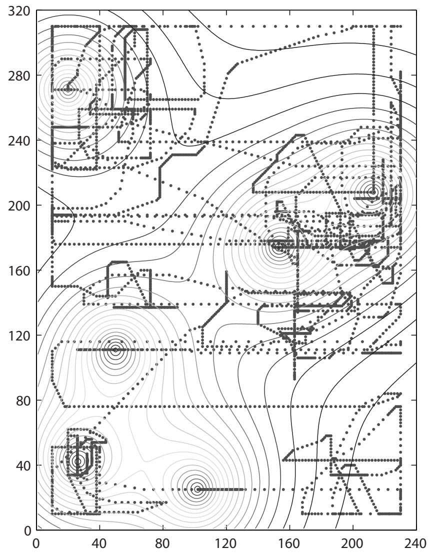
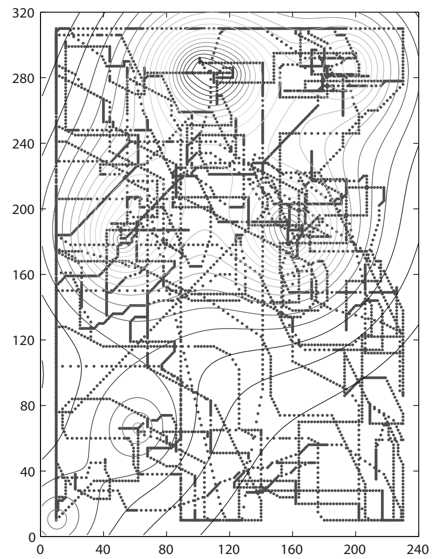


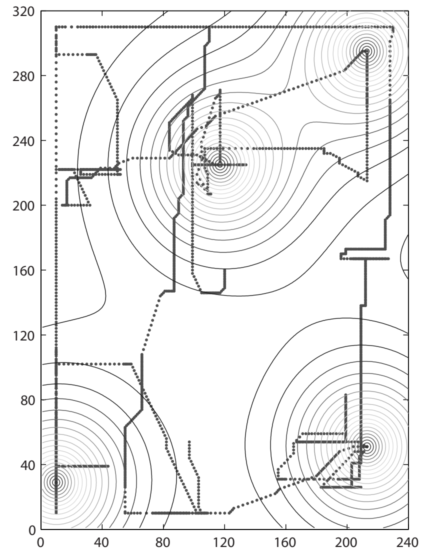
Figure 8. Examples of few search patterns in different conditions



a) Example of quadrant search: Subject 9, "Doppler-no vibration"



b) Example of sector search: Subject 1, "no Doppler-no vibration"



c) Example of no formulaic search: Subject 7, "Doppler-no vibration"

was the sweep search (76%). The next most common was the no distinguishable pattern (14%), followed by quadrant (6%), and sector (4%). Participants' audio and exploration density plots show "Doppler-no vibration" has the least covered space with 34.5%, and the rest have similar percentage of coverage, 37.6%.

Chosen Songs

The accuracy relative to the number of song chosen is another factor in improving audio interfaces. Because the type of songs may affect the perception of distortion due to the Doppler effect, and affect the users' ability to recognise and locate them. We measured the number of audio sources

Figure 9. Mean distance in pixels from target in different tasks

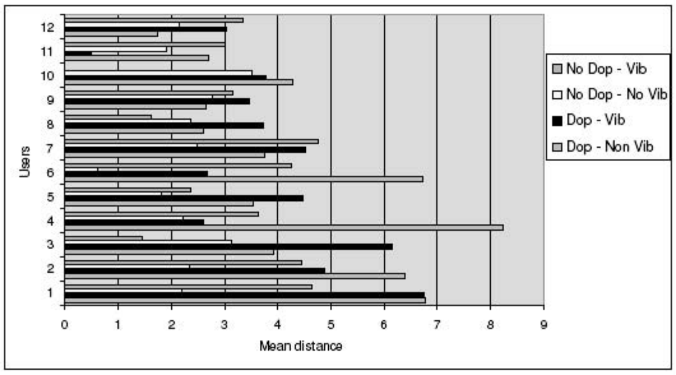


Figure 10. Count of most accurately chosen songs in different conditions for all users

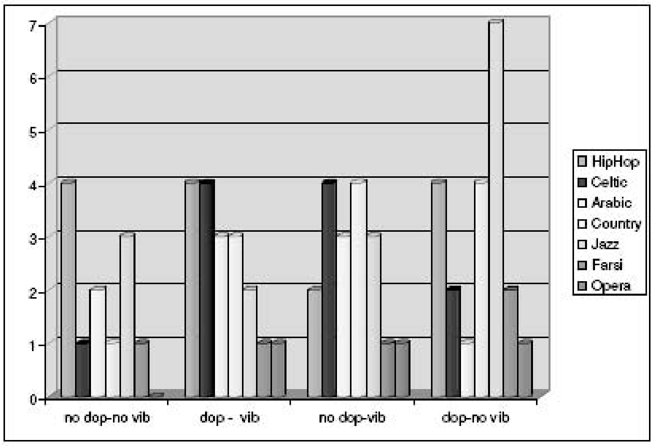


Figure 11. Mean distance (pixels) of selected songs in all conditions for all users.

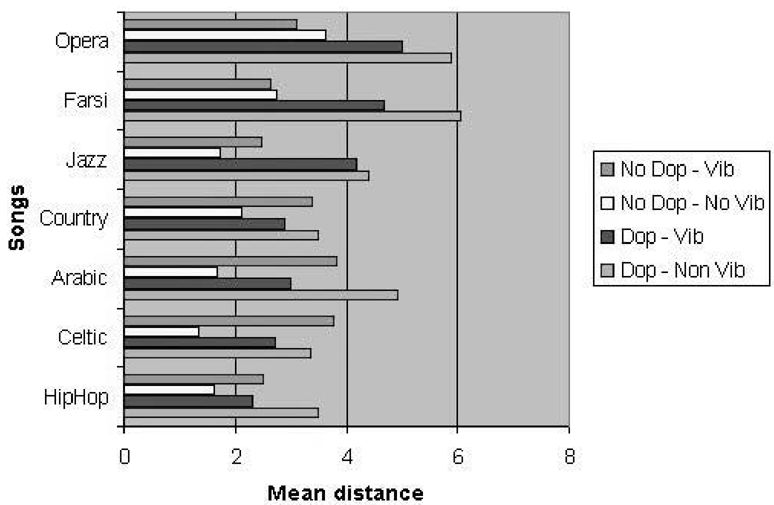


Table 2. Accuracy scores for audio sources in the first experiment

-	no dop- no vib	Dop-vib	no dop-vib	dop-no vib
Hip-Hop	4	4	2	4
Celtic	1	4	4	2
Arabic	2	3	3	1
Country	1	3	4	4
Jazz	3	2	2	7
Farsi	1	1	1	2
Opera	0	1	1	1

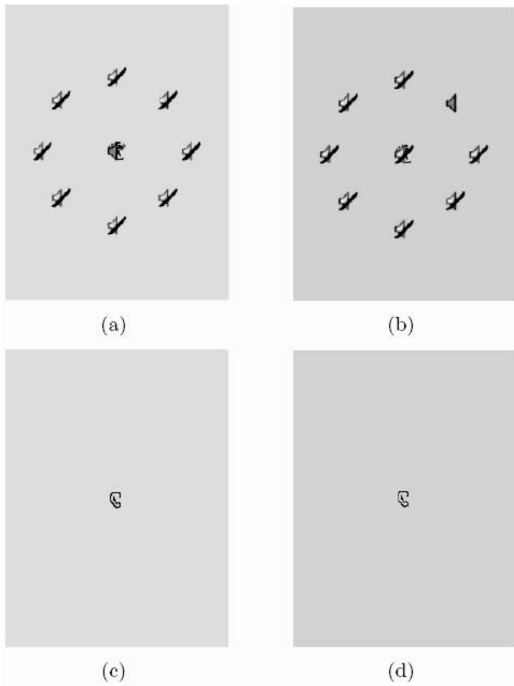
participants have selected. The mean accuracy for each of these sources has been summarised in Table 2. Figure 10 shows the mean accuracy count of songs in all conditions for all participants. This result is based on the number of times each source was selected with the smallest distance to the target in each condition. There is a large amount of variability in the results. Jazz music was selected more than others, on average, but Hip-hop music was chosen more accurately in the "no Doppler-no vibration" condition. Figure 11 shows mean error for songs in all conditions. In general "no Doppler-no vibration" has the lowest error among others and "Doppler-no vibration" has the highest error. Farsi and Arabic sources had high mean and maximum errors in the Doppler case.

Discussions

Post hoc examinations of the cursor's trace in this experiment showed that the subjects tended to use the same technique regardless of the sounds they heard and the audio condition. Six subjects (50%) used one search technique exclusively. Of these six subjects, five used the sweep technique, one used no distinguishable pattern exclusively. Another six subjects (50%) used the same technique in two out of the four conditions. This consistency in the application of a searching technique has several notable points. First, the same technique was employed regardless of the sound treatment. This would indicate that the subjects brought with them a technique that was not altered by the change in the treatments used in the auditory interface. The subjects were given no experimental feedback that might prompt them to change their search pattern to one that might be more effective. Left to their own means, the subjects tended to continue with the application of the search pattern with which they felt most comfortable. Second, the most common type of search pattern (sweep search) was also the least effective, given that the target in all four conditions was located towards the interior of the information space. In these cases, the subject was less likely to notice a change in the sounds they were hearing because of the low intensity of the sounds generated at the borders of the information space. Because they typically did not explore the interior, they would not hear the more intense sounds that might lead them to the target. In conditions with vibration feedback sweep search is combined with circular movements around the vibration source (Figures 6(a), 7(a) and 7(b)) and has led the users to the target. This suggests that the vibration was more important for the users in locating a target, and whenever they felt they were close to the song, they looked for the vibration source before clicking; so, feeling a vibration source meant they were at the centre of the audio source. This might also explain the fact that errors were not smaller, as the user may often have selected the location as soon as vibration was perceived, at the edge of the circle, rather than at the centre of the target itself.

The "no formulaic search" is the least thorough of the systematic techniques. Even though there was essentially no effort involved in exploring more thoroughly by applying a different search pattern, the subjects tended to use the no distinguishable search pattern. This could be accounted

Figure 12. The state space in experiment II and corresponding angles. Top pictures show the training application and bottom ones show the main application. Left pictures show the screen before pressing the button for Jazz music and right ones show the screen after pressing the button (colour changes and covered speakers are indicated in training application).



for by assumptions the subjects made about the nature of the information space. It would seem that some subjects took the experimental task seriously by systematically exploring the information space. Other subjects did not seem to be interested in exploration, but instead made a quick "stab" in the general direction of the high point. The case could be made that those subjects who explored liked the sonic interface, and those who did not explore did not like the interface. It may well be the case that auditory display is not for everyone. Some will like it and make use of it, others will not.

The results show that the mean distance from the selected position to the target in "no Doppler-no vibration" is less than other experiments (Figure 9). The extra clicks and navigation activities in the cursor trajectories for Doppler might be an effect of the extra sensitivity of the feedback to movement, which makes the users explore by varying their velocity vector. Variability in localisation accuracy is greater with the Doppler effect for the Farsi and Arabic sources, suggesting that for the mainly western European participants, their poorer familiarity with these sources made the distortions introduced by the Doppler effect more difficult to perceive. Opera also had larger errors, again suggesting that less familiarity with the target sources can affect the usefulness of this approach. The large number of falsely placed points for the Doppler method might be because of the amplification involved in moving towards something and potentially, frequency and speed of sound, which makes people feel they are getting a stronger response, and they over-interpret the quickened signal, believing they are already at the point - a common cited risk associated with quickened displays (Poulton, 1974).

Experiment II

Twenty-four paid participants (8 male, 16 female), all sighted with normal or corrected vision and normal hearing with mean age 24, were recruited through sign-up sheets in Glasgow University Psychology department and through e-mail. Two of the participants were left handed. Three participants had experience using a Pocket PC, and of these, one had frequent use. None had experience using an accelerometer as an input device.

Design

Given that the results of the pilot study showed we had possible confounding factors, that is, four different audio conditions, haptic feedback, few subjects, unfamiliar songs, and random located targets, the next experiment reduced the number of independent variables. This resulted in three different types of audio feedback without haptic feedback:

1. Doppler feedback
2. Derivative volume adaptation
3. No quickening

There were eight possible audio sources (targets) arranged in a circle (radius = 100 pixels) around the centre at 45° intervals (Figure 12). The audio feedback at the centre was jazz music, which played continuously for all conditions. When an outside target was to be located, the audio feedback was "Hotel California" played in a loop. The audio source to be located alternated between the centre (jazz music) and one of the outside targets ("Hotel California"), and always began with the centre target. The audio sources around the outside were presented in a random order, twice for each target for the training session (16 trials in all), and five times for each target for the experimental trial (40 trials altogether). Once one target had been located, a button was pressed. For each key pressing, there was a screen colour change and a short "beep" sound. Audio was noticeable within a radius of 90 pixels from sources (3s) in conditions 1 and 2, but audio was noticeable just within a radius of 15 pixels from the sources in condition 3, and there was no feedback at any other locations in this condition. Each participant was tested individually, and participants were told to commence the training session when they felt ready. After the training session, it was ensured that they understood the procedure fully and that they felt comfortable using the equipment. They were given a break between the training and experimental sessions if they wanted one.

Visualisation

In addition to exploration and audio density plots and distance to the target, we used another visualisation method that measures the orientation of each target with respect to the centre point, showing which angles in the state space have had the most accurate data in selecting targets. This measurement is important in this experiment to see whether the orientation of audio sources has any effect on the targeting task. Results in this experiment were analysed using a *GLM ANOVA* test.

Results

Proportion of Distance to the Target

Figure 13 shows the box plot of medians, means, and measures of spread of the distance between the audio sources and the position selected for each of the three audio feedbacks. The red triangle indicates the mean and the blue square indicates the median (50% of the observations lie below this line). The top of each box indicates the upper quartile (75% of the observations lie below this point) and the bottom indicates the lower quartile (25% of the observations lie below this point). The tops of the lines above the boxes indicate the highest observation, and the bottom of the lower line indicates the lowest observation. The blue stars indicate outliers: those observations that differ significantly from the mean. Figures 13 and 14 show a difference in the average distances between the actual audio source and the target selected (accuracy) for the three audio conditions. The most accurate target selection

Figure 13. Boxplot of distance versus audio conditions and angles

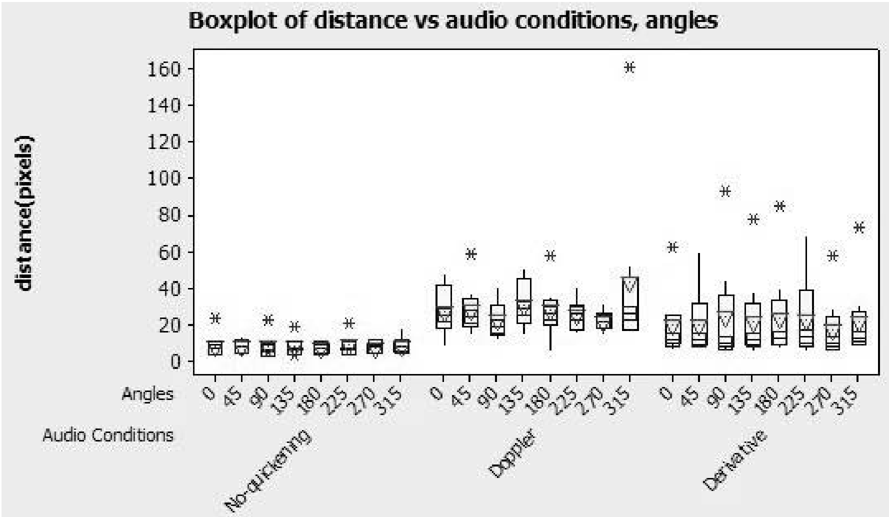
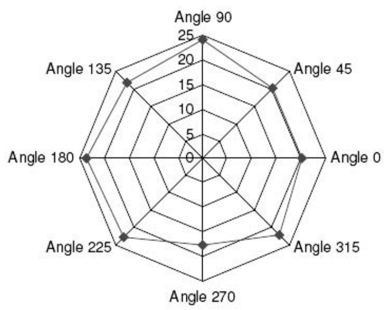
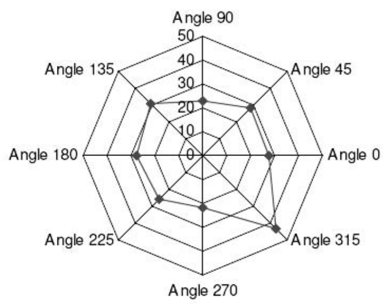


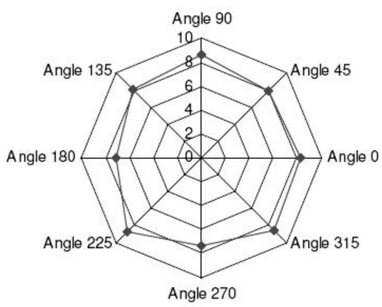
Figure 14. Average of position error in pixels for all participants in 3 audio conditions



(a) Average of position error in the Derivative condition for all users



(b) Average of position error in the Doppler condition for all users



(c) Average of position error in the No-Quickening condition for all users occurred in the no-quickening condition, and there was little difference between the levels of accuracy for each orientation. The spread for no quickening was very small, and the five outliers are not very far away from the median, suggesting that, overall, most people in this condition took approximately the same length of time to select a target. The derivative condition takes longer overall than the no-quickening condition. The spread is larger (largest of all three conditions) and the outliers further away from the median than in the no-quickening condition. The participants in the Doppler condition took longer on average to select the targets, compared with those in the other two conditions. The spread in the Doppler condition is smaller than in the derivative condition, but larger than in the no-quickening condition. This shows that angles 90° and 270° have higher accuracy. A *GLM ANOVA* analysis found that there was a significant effect of audio type on the distance ($F(2,21)=4.345$; $p<0.05$). There were, however, no significant differences between the eight audio orientations. Post-hoc analysis using a *Tukey* test showed that there was a significant difference between the estimated mean distances of the selected targets away from the audio sources in two of the three audio conditions. It was found that the mean distances were not significantly different between two quickening feedbacks and nonquickening condition ($p<0.087$). The *Tukey* test showed that there was no significant difference between the estimated mean distances of the selected targets away from the audio sources in the derivative condition. Running this test on Doppler results revealed that a significant difference ($p<0.025$) between the estimated mean distances of the two selected targets away from the eight audio orientations (90° and 270°). In the no-quickening condition, the *Tukey* test showed that there is no significant difference among the estimated mean distances of the selected targets away from the audio sources.

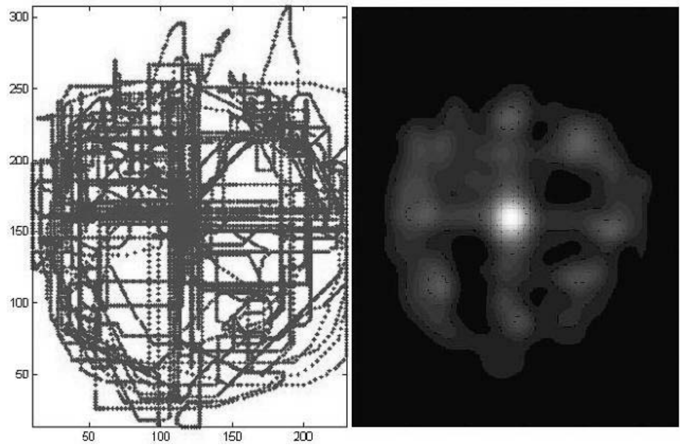
Search Patterns and Covered Space

In this experiment, audio targets had fixed positions in the state space so the observed search patterns were different than those we found in the pilot study. Twenty-four total patterns were analysed in this experiment. From the training sessions, subjects knew the approximate location of the audio sources. In 100% of search patterns in the no-quickening condition, the subjects moved to the edge of a circle in the size of the actual radius of the points and started circling to find the active target. In the Doppler condition, 87% of subjects could guess in which direction the target was located and after doing a few back and forth movements, they landed on the target. Subjects, therefore, followed a sector search pattern. Search patterns of subjects who worked with derivative volume adaptation were mixtures of the patterns of the no-quickening and Doppler conditions. Figure 15 shows some of the subjects' trajectories and density plots in different conditions. Figure 16 shows the percentage of the screen covered by participants' movement in three conditions. In the derivative condition, the top-left sections (90°-180°) were explored more than other parts. In the Doppler condition, the top-right (0°-90°) sections were popular to explore, and in "no quickening" there was no significant difference in the sections covered by participants' movement and all of the partitions were explored equally. These plots show the Doppler condition had the most covered space with 44.5% and the rest had a similar percentage of coverage, 39%.

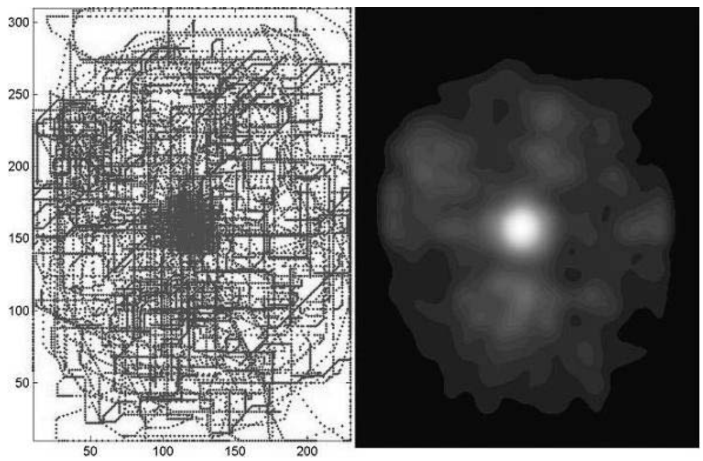
Discussions

In this experiment, it was found that there was an effect of audio condition on the level of accuracy. When the feedback was no quickening, participants were more accurate than when the feedback was Doppler or derivative. This is due to the fact that with the no-quickening condition, the only time that audio feedback is heard is when the cursor is directly over the target, and the difference between hearing and not hearing audio feedback is larger than the difference between hearing different levels of audio feedback. There was also found to be no effect of angle. The level of accuracy was the same, irrespective of the orientation of the target.

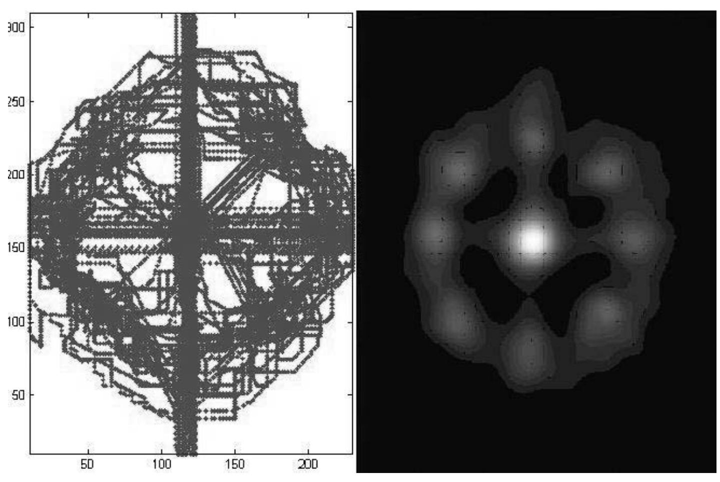
Figure 15. Trajectories of different subjects in 3 audio conditions



a) Trajectory and density plot of subject 4 in Derivative

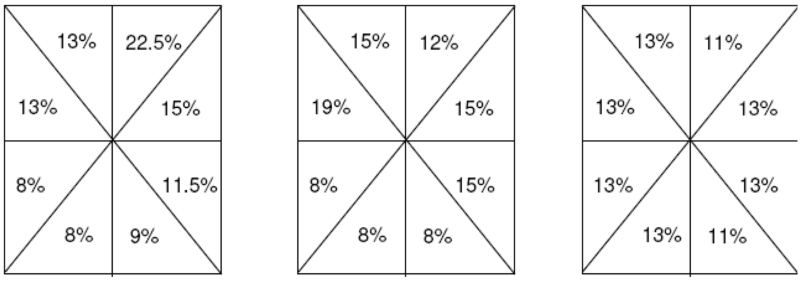


b) Trajectory and density plot of subject 4 in Dopler



c) Trajectory and density plot of subject 4 in no quickening

Figure 16. Percentage of the screen covered by users' movement in different conditions. The variance of the coverage in the derivative condition is 0.0023, in the Doppler condition is 0.0017 and in the no-quickening condition is 8.5714e-005.



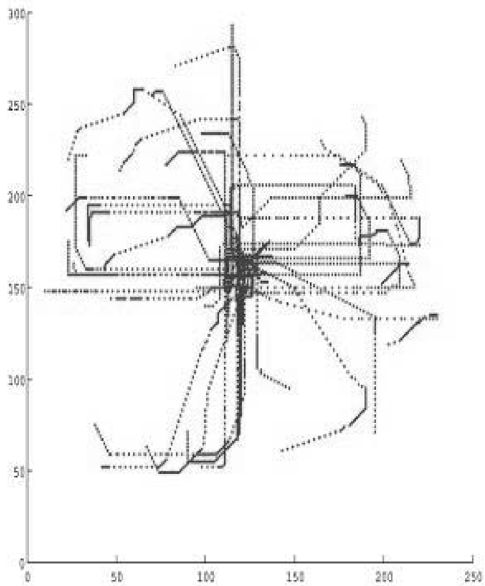
(a) Derivative (b) Doppler (c) No-Quickening

Many participants reported that sometimes they would just "land" on the audio source by chance, and at other times they would search for a long time and still not feel they had located the point accurately. This was an especially common complaint by participants in the Doppler condition. Disorganised search patterns observed in the Doppler condition, for instance in Figure 15(b), may correspond with the claim made by some participants that by the time they had established where the target was through the audio feedback, they had already passed the audio source, and had to go back to it. We plotted users' trajectories when they had moved from the centre of the screen to any active audio source around the centre. Figure 17(b) and other users' trajectories in the Doppler condition highlight that in the first moments after audio source activation, the subjects could guess the direction (left or right side of the space) and approximate position of the target and consequentially, moved towards the target correctly, but it was difficult to establish a correct target acquisition, and they made some back and forth or up-down movements to land on the target, which is compatible with the observed sector search pattern. This is shown more clearly in Figure 18(a-left) and the user's trajectories as time series in Figure 18 (b-left) in an individual target acquisition task when the user has moved from the centre to the target in angle 45°, which has been activated.

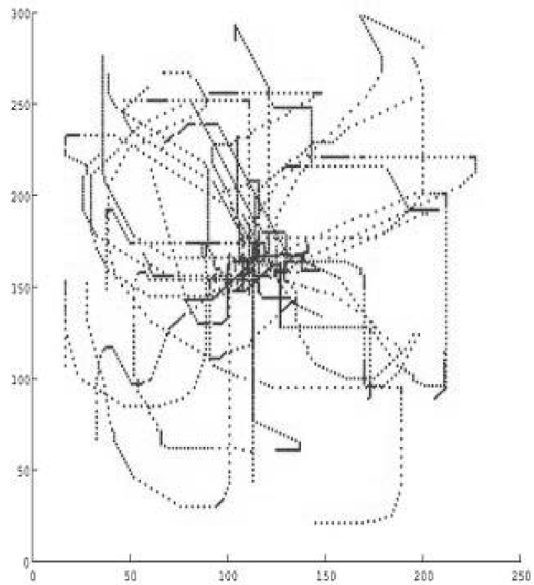
Figures 15(a) and 17(a) show a trajectory that is fairly typical for most participants in the derivative condition. It can be seen that the trajectory is far more ordered, with participants moving in the horizontal and vertical directions (in the directions of 0°, 90°, 180°, and 270°), more so than in the Doppler condition. It becomes obvious that participants moved in a circular motion that they learnt during the training session, far more so than those in the Doppler condition. From this, they have established that the audio sources in the experimental session were also arranged in a circle. This suggests that they were not necessarily using only audio feedback, but also prior knowledge about the probable locations of the audio sources. Since this circle is not as clearly defined in the Doppler trajectories, it suggests that participants in the Doppler condition were using predictive information, but were also less able to control their movements efficiently. There is a risk that any significant effects were masked by prior knowledge of the way the audio sources were arranged, and visual feedback has affected the users' behaviour in exploring the audio space.

Figures 17(a) and 19(a-left) provide a clearer picture of the users' browsing behaviour in this condition. As a result of the first impressions that the users have received from the volume of the audio source, they have chosen vertical or horizontal directions. Whenever they have not found the target in these directions, for instance in an individual target acquisition in angle 45° presented in Figure

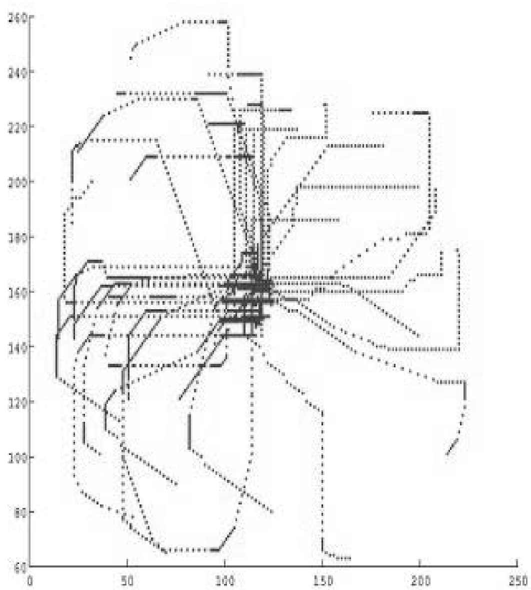
Figure 17. Trajectories of different subjects in 3 audio conditions when they have moved from the centre to outlying active audio sources



a) Trajectory and density plot of subject 1 in Derivative case

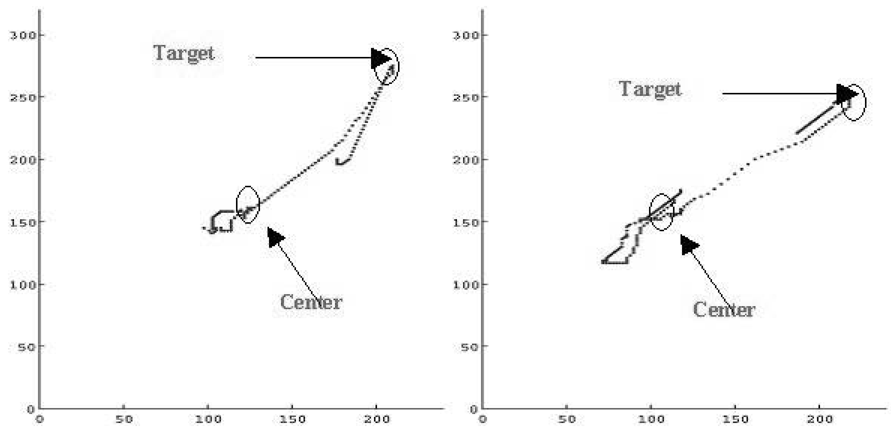


b) Trajectory of subject 4 in Doppler case

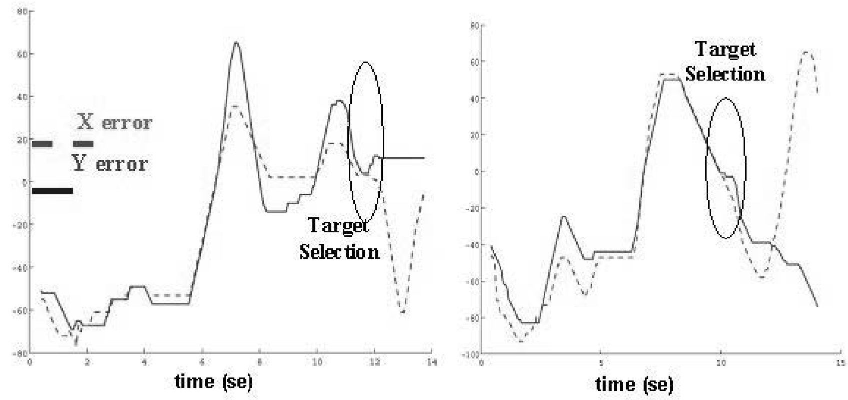


c) Trajectory of subject 4 in no quickening case

Figure 18. (a) One of the participants' trajectories in the Doppler condition with and without predictive feedback in an individual target acquisition task when the user has moved from the centre to an activated target in angle 45°. (b) The time series of the participant's X and Y position error in the same task.



(a) An individual target acquisition task in the Doppler condition. (left) standard feedback, (right) predictive feedback



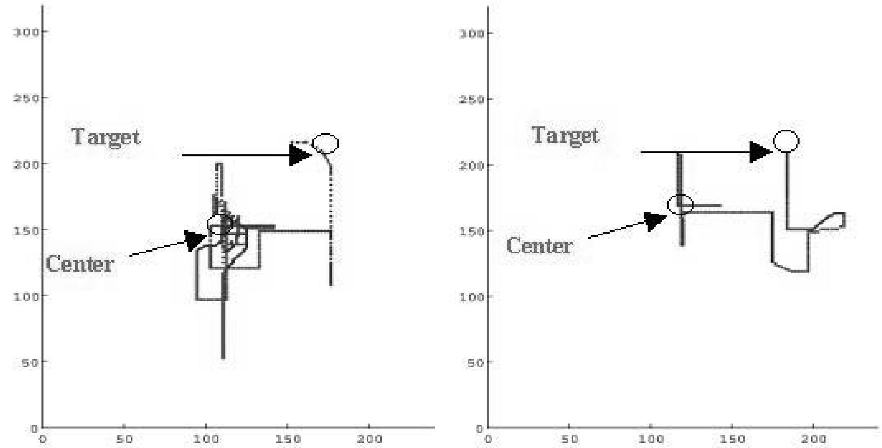
(b) The time series of one of the participants' position error in X and Y axis in the target acquisition task shown above, in the Doppler condition. (left) standard feedback, (right) predictive feedback

19, they have moved around the circle using prior knowledge of the landscape. Figure 15(c) shows a more pronounced circle and cross-shape for the no-quickening condition. Since the participants in the no-quickening condition were presented with no aural feedback except when directly over the target, it is, most likely that they were relying on the circular target distribution previously seen in the training. This led to a systematic search strategy, less "browsing around" because of the lack of predictive ability without quickening. All users who participated in this condition claimed that this was not an exciting method of exploring the auditory space.

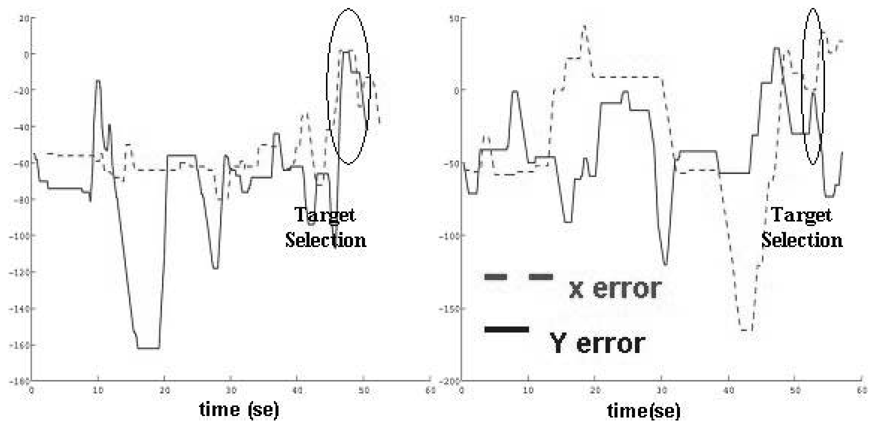
Human Operator Modeling

In continuous control tasks, for instance browsing and finding audio targets in the audio space, the human operator can be modeled using the tools of

Figure 19. (a) One of the participants' trajectories in the derivative condition with and without predictive feedback, in an individual target acquisition task when the user has moved from the centre to an activated target in angle 45°. (b) The time series of the participant's X and Y position error in the same task.



(a) An individual target acquisition task in the Derivative condition (left) standard feedback, (right) predictive feedback



(b) The time series of one of the participants' position error in X and Y axis in the target acquisition shown above, in the Derivative condition (left) standard feedback, (right) predictive feedback

manual control theory. Quantitative models of the human operator may provide predictions and insights into basic properties of human performance in human-machine interaction, and the ability to derive transfer function for human operators means they can be directly used in human-machine systems interaction design (Jagacinski & Flach, 2003).

Human interaction with the computing systems is two-way. The user and machine form a closed-loop system, where s/he issues commands through the system's input channels and receives results fed back on output channels. Subsequent input depends on the latest output. The performance is adversely affected when the feedback is subject to delay or lag (Jagacinski & Flach, 2003; MacKenzie & Ware, 1993).

In some genres of interactive systems, which rely heavily on the tracking of hand, head, and/or body motion in a simulated environment, the pretence of reality requires a tight coupling between the user's view and hearing of the environment and the actions, usually hand, head, and body motions, that set the view and hearing. When changes in the environment lag behind input motions, the loss of fidelity is dramatic (MacKenzie & Ware, 1993). In an extension to the previous experiments, we carried out a primary investigation to measure and model the speed, accuracy, and bandwidth of human motor-sensory performance in interactive tasks subject to lag.

Manual control theory suggests that in a simple tracking task the human operator can be modeled via a transfer function that consists of a gain, a lag (an integrator at higher frequencies) and a time delay (Jagacinski & Flach, 2003; Poulton, 1974).

$$Y_h(j\omega) = \frac{Ke^{-j\omega\tau}}{j\omega} \qquad (3)$$

The gain $\boldsymbol{K}$ is a scaling factor that influences the bandwidth of the control system. The time delay t reflects human reaction time. In simple tracking tasks, the range of the time delay is between 20 to 150 ms, which overlaps with measures of reaction time in response to continuous stimuli. If $\boldsymbol{K}$ is low, the system will respond very sluggishly, moving only slowly towards the target signal. Conversely, if $\boldsymbol{K}$ is high then the system is likely to overshoot, requiring adjustment in the opposite direction, which itself may overshoot, leading to oscillation. However, humans adjust their gain to compensate increases or decreases in plant gain. For example, pilots change their behaviour when they switch from Boeing 747 (heavy) to an aerobatic airplane so the total open-loop gain remains constant. So if Y_p represents the plant transfer function and Y_h represents the human transfer function, then:

$$Y_h(j\omega)Y_p(j\omega) = const \qquad (4)$$

The delay τ can also contribute to this behaviour; a high delay makes oscillatory behaviour much more likely (refer to time and delay section discussed earlier). The lag suggests that the human tracker has a low pass characteristic, that is, the human responds to low-frequency components of errors and ignores (or filters out) the high-frequency components of error (MacKenzie & Ware, 1993).

Using the platform in the previous experiment, we did a preliminary investigation to measure and model the accuracy, and bandwidth of human motor-sensory performance in interactive tasks subject to lag. We kept the same format of the second experiment but instead of providing audio feedback to the user's current position, we provided feedback to the user's predicted position, calculated according to equation (3):

$$X_{t+\tau} = X_t + V\tau \qquad (5)$$

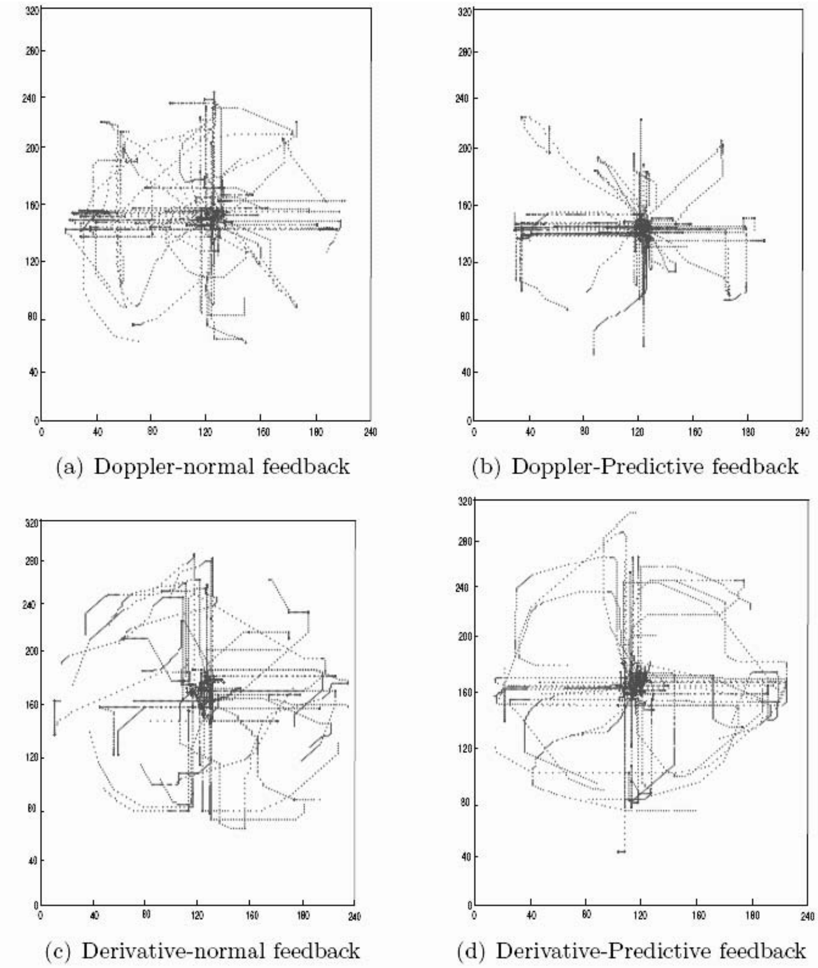
The volume of the audio source, which provides feedback about the target's position, is a function of the user's current velocity and position. Here, X_{t+1} and X_t are the user's position at time $t + \tau$ and t or current position and next possible position respectively. τ is the human's time-delay or reflection time, which becomes our "prediction horizon" for the predictive model, and V is the user's speed of motion in the audio space. This has the effect that, as the user moves toward the target, s/he feels him/herself in the position predicted to be reached at time $t + \tau$.

Design

In the second experimental setup, we added a smooth drop-off in the time horizon of the prediction as the target was approached. The falloff began at radius 15 pixels, and once the user was within 5 pixels of the source, the feedback reverted to standard feedback with no predictive element.

In a pilot study, we ran the application for three participants familiar with the Pocket PCs and accelerometer. Neither felt any difference in the two derivative conditions, with and without prediction. In Figures 19(a) and 20(b), we see one of the participant's trajectories when he has moved from the centre to any active target. Providing feedback to the user's future position in the derivative condition has not much changed the user's exploratory behaviour. But users reported a great difference between Doppler with the pre-

Figure 20. The trajectories of 2 subjects in the Doppler and derivative conditions, without and with predictive feedback



(a) Doppler-normal feedback

(b) Doppler-Predictive feedback

(c) Derivative-normal feedback

(d) Derivative-Predictive feedback

dictive model-based feedback and Doppler with no predictive feedback. They said it felt they were able to acquire the direction of the audio source more quickly in the predictive case, but that it was more difficult to land on the source. In the standard model with no predictive element they felt it was slower to find the direction, but easier to land. Despite their perceptions, the trajectories in Figures 18(a-right) and 20(a) suggest the opposite case, that they performed fewer oscillatory movements around the target in the predictive model, compared to the standard case.

Prediction in the Doppler case allowed the users to converge more rapidly and directly to the target, but it seemed less helpful very close to the target. In the derivative condition, predictive feedback seems to have smoothed the behaviour, but has not improved the initial target localisation. These preliminary explorations suggest that a more detailed investigation of incorporating the predictive element in the feedback system would be of interest.

CONCLUSION AND SUMMARY

This work presents initial experimental results exploring the use of quickened audio displays for localisation and selection based on tilt control of mobile devices. The experiments provided useful exploratory information about how users navigate in such environments, and highlighted some benefits and disadvantages of each of the display options investigated. Users used a search method, which they felt more comfortable with for browsing the space regardless of the sound treatment. Vibration was clearly perceived by users, but led them to spend more time circling around targets.

Average results in the pilot study on the metrics used suggest that participants were more accurate in target selection in the "no Doppler-no vibration" than other conditions. The results do not suggest that the use of Doppler or vibration brought consistently improved accuracy, but some people did very well with Doppler, and most stated that they found the vibration feedback useful. Longer studies might show different use in real-life tasks once users had familiarised themselves with the system.

The main study represented a more focused investigation, with fewer confounding factors. We increased the number of participants, placed all the targets at equal distances from the starting point (centre of the screen), did not include haptic feedback, chose western pop music that was familiar to all users, and allocated more time for allowing users to learn how to use the specific interface, which was new to all. We investigated whether quickening was more useful to users searching for targets in state-space than no quickening audio feedback. We also investigated if there was any advantage of using Doppler feedback over derivative, and if there was an effect of orientation in either Doppler, derivative, or no quickening; therefore, to find out if the results of the first experiment could have been masked by an interaction with the orientation of the targets. It was also found that there was no effect of the angle at which the audio source was located.

In a preliminary investigation to better understand the results and to guide future work, we performed an exploratory experiment with predictive model-based feedback. The model is based on human operator modeling in continuous tracking tasks, and it could take human response delays and lags into account (not considered in this work). Using this predictive model, we could improve the users' performance in the Doppler condition, and reduce their overshoots during landing on the target just by providing audio feedback about the user's predicted position instead of their current position. This suggests further research to investigate the benefits of explicitly incorporating models of human behaviour in the design of feedback methods.

Outlook for Mobile Interface Designers

These results are a useful starting point for further investigation into the types of feedback that are most useful and informative in assisting users of a tilt-controlled mobile device with multimodal feedback. Some of the visualisation tools used will be useful for other designers, but the work also gives an indication of the difficulty of designing experiments that test aspects of low-level perception of multimodal displays, without confounding factors from prior knowledge influencing the results. The experiments also show the need for longitudinal studies. As in early exploration of novel interfaces, much observed behaviour is related to the user exploring the novel interface, and might not be a reliable indicator of typical practiced behaviour. Supporting the design of interaction in mobile devices with multimodal interfaces is a key challenge in mobile HCI. We believe that further development of the model-based prediction techniques we have begun to explore in this chapter will not only give us a better understanding of typical user behaviour, but will provide a promising, scientific basis to support designers in creating more useable systems in a wide range of novel settings, with a range of sensors and displays.

ACKNOWLEDGMENT

The authors gratefully acknowledge the support of IRCSET BRG SC/2003/271 project *Continuous Gestural Interaction with Mobile devices*, HEA project *Body Space*, and SFI grant 00/PI.1/C067, the IST Programme of the European Commission, under PASCAL Network of Excellence, IST 2002-506778. This publication only reflects the views of the authors.

REFERENCES

Birmingham, H. P., & Taylor, F. V. (1954). A design philosophy for man-machine control systems. In *Proceedings of the Institute of Radio Engineers* (pp. 1748-1758).

Blattner, M., Papp, A., & Glinert, E. (1992). Sonic enhancements of two dimensional graphic displays. In *Proceedings of International Conference on Auditory Display, ICAD'92* (pp. 447-470). Santa Fe, NM.

Bootsma, R. J., Fernandez, L., & Mottet, D. (2004). Behind Fitts' law: Kinematic patterns in goal-directed movements. *International Journal of Human-Computer studies, IJHCS*, *61*(1), 811-821.

Bregman, A. S. (1990). *Auditory scene analysis: The perceptual organisation of sound.* Cambridge, MA: MIT Press.

Brewster, S. A. (1997). Using non-speech sound to overcome information overload. *Special Issue on Multimedia Displays* (pp. 179-189).

Brewster, S. A. (2002). Overcoming the lack of screen space on mobile computers. In *Proceedings of Personal and Ubiquitous Computing* (pp. 188-205).

Brewster, S. A., & Murray, R. (1998). Presenting dynamic information on mobile computers. *Personal Technologies* (pp. 209-212).

Chion, M. (1994). *Audio-vision*. New York: Columbia University Press.

Cook, P. R. (2002). *Real sound synthesis for interactive applications.* Peters, A K, Limited.

Cook, P. R., & Lakatos, S. (2003). Using Dsp-based parametric physical synthesis models to study human sound perception. *IEEE Workshop, Applications of Signal Processing to Audio and Acoustic*s (pp. 75-78).

Crossman, E. R. F. W., & Goodeve, P. J. (1983). Feedback control of hand-movement and Fitts' law. *Quarterly Journal of Experimental Psychology, (Original work presented at the meeting of the Experimental Psychology Society, Oxford, England, July 1963), 35A*, 251-278.

Doherty, G., & Massink, M. (1999). Continuous Interaction and Human Control. In *Proceedings of the XVIII European Annual Conference on Human Decision Making and Manual Control* (pp. 80-96). Loughborough: Group-D Publications.

Dong, L., Watters, C., & Duffy, J. (2005). Comparing two one-handed access methods on a PDA. In *Proceedings of the MobileHCI'05* (pp. 291-295). Salzburg, Austria.

Eslambolchilar, P., Crossan, A., & Murray-Smith, R. (2004). *Model based target sonification on mobile devices.* Paper presented at the Interactive Sonification Workshop, ISON, Bielefeld University, Germany.

Faconti, G., & Massink, M. (2001). Continuous interaction with computers: Issues and requirements. In *Proceedings of Universal Access in HCI, Universal Access in HCI - HCI International 2001* (pp. 301-304), New Orleans.

Fallman, D. (2002a). An interface with weight: Taking interaction by tilt beyond disembodied metaphors. *4th International Symposium on Mobile Human-Computer Interaction* (pp. 291-295).

Fallman, D. (2002b). Wear, point and tilt. In *Proceedings of the Conference on Designing Interactive Systems: Processes, practices, methods, and techniques* (pp. 293-302).

Fitts, P. M. (1954). The information capacity of the human motor system in controlling the amplitude of movement. *Journal of Experimental Psychology, 47*(1), 381-391.

FMOD. (2004). *Music and sound effect system.*

GAPIDraw. (2004). *Graphics API Draw.*

Gaver, W., Smith, R., & O'Shea, T. (1991). Effective sounds in complex systems:The ARKola simulation. In *Proceedings of the ACM Conference on Human Factors in Computing Systems CHI'91* (pp. 471-498).

Gaver, W. W. (1989). *The sonic finder* (Vol. 4). Elsevier Science Publishing.

Gaver, W. W. (1997). *Auditory interfaces* (2nd ed.): *Handbook of human-computer interaction*.

Harrison, B., & Fishkin, K. P. (1998). Squeeze me, hold me, tilt me! An exploration of manipulative user interfaces. In *Proceedings of the ACM Conference on Human Factors in Computing Systems, CHI'98* (pp. 17-24), Los Angeles.

Hermann, T., Hansen, M. H., & Ritter, H. (2000). Principal curve sonification. In *Proceedings of International Conference on Auditory Displays, ICAD'00* (pp. 81-86).

Hermann, T., & Ritter, H. (1999). Listen to your data: Model-based sonification for data analysis. *Intelligent Computing and Multimedia Systems* (pp. 189-194). Baden-Baden, Germany.

Hinckley, K., Pierce, J., Horvitz, E., & Sinclair, M. (2005). Foreground and background interaction with sensor-enhanced mobile devices. *ACM Transactions on Computer-Human Interaction (TOCHI)* (pp. 31-52).

Hoffmann, E. R. (1991). Capture of moving targets: A modification of Fitts' law. *Ergonomics, 34*(2), 211-220.

Holmes, J. (2005). Interacting with an information space using sound: Accuracy and patterns. In *Proceedings of ICAD 05-Eleventh Meeting of the International Conference on Auditory Display* (pp. 69-76). Limerick, Ireland.

Jagacinski, R. J., & Flach, J. M. (2003). *Control theory for humans: Quantitative approaches to modeling performance*. Lawrence Erlbaum Associates, Inc.

Jenison, R. L. (1997). On acoustic information for motion. *Ecological Psychology, 9*(2), 131-151.

Johnson, C., Brewster, S. A., Leplatre, G., & Crease, M. G. (1998). Using non-speech sounds in mobile computing devices. *First Workshop on Human Computer Interaction with Mobile Devices*, Department of Computing Science, University of Glasgow, Glasgow, UK (pp. 26-29).

Kramer, G., Walker, B., Bonebright, T., Cook, P., Flowers, J., Miner, N., et al. (1999). *Sonification report: Status of the field and research agenda*. Report prepared for the National Science Foundation by members of the International Community for Auditory Display.

Langolf, G. D., Chaffin, D. B., & Foulke, J. A. (1976). An investigation of Fitts' Law using a wide range of movement amplitudes. *Journal of Motor Behaviour, 8*, 113-128.

MacKenzie, I. S., & Ware, C. (1993). Lag as a determinant of human performance in interactive systems. In *Proceedings of the ACM Conference on Human Factors in Computing Systems - INTERCHI '93* (pp. 488-493). Amsterdam.

Mackinlay, J. D., Robertson, G. G., & Card, C. K. (1991). The perspective wall: Detail and context smoothly integrated. In *Proceedings of the ACM Conference on Human Factors in Computing Systems, CHI'91* (pp. 173-179). New Orleans, LA.

Oakley, I., Ängeslevä, J., Hughes, S., & O'Modhrain, S. (2004). Tilt and feel: Scrolling with vibrotactile display. In *Proceedings of Eurohaptics 2004,* Munich, Germany.

Partridge, K., Chatterjee, S., Sazawal, V., Borriello, G., & Want, R. (2002). TiltType: Accelerometer-supported text entry for very small devices. *UIST'02: Proceedings of the 15th Annual ACM Symposium on User Interface Software and Technology* (pp. 201-204). Paris.

Patrol, C. A. P.-R. A. (1999). *Mission observer course manual*. Civil Air Patrol.

Poulton, E. C. (1974). *Tracking skill and manual contro*. Academic press.

Rekimoto, J. (1996). Tilting operations for small screen interfaces. *UIST'96: Proceedings of the*

9th Annual ACM Symposium on User Interface Software and Technology (pp. 167-168). Seattle, WA.

Rinott, M. (2004). Sonified interactions with mobile devices. In *Proceedings of the International Workshop on Interactive Sonification*, Bielefeld, Germany.

Sazawal, V., Want, R., & Borriello, G. (2002). The unigesture approach, One-handed text entry for small devices. In *Proceedings of Mobile Human-Computer Interaction - Mobile HCI'02* (pp. 256-270).

Smith, D. R., & Walker, B. N. (2005). Effects of auditory context cues and training on performance of a point estimation sonification task. *Applied Cognitive Psychology, 19*(8), 1065-1087.

Smith, N., & Schmuckler, M. A. (2004). The perception of tonal structure by the differentiation and organization of pitches. *Journal of Experimental Psychology: Human Perception and Performance, 30,* 268-286.

Walker, B. N., & Lindsay, J. (2006). Navigation performance with a virtual auditory display: Effects of beacon sound, capture radius, and practice. *Human Factors, 48*(2), 265-278.

Wigdor, D., & Balakrishnan, R. (2003). TiltText: Using tilt for text input to mobile phones. *UIST'03: Proceedings of the 16th annual ACM symposium on User interface software and technology* (pp. 81-90). Vancouver, Canada.

Williamson, J., & Murray-Smith, R. (2004). Pointing without a pointer. In *Proceedings of the ACM Conference on Human Factors in Computing Systems, CHI'04* (pp. 1407-1410). Vienna, Austria.

Williamson, J., Strachan, S., & Murray-Smith, R. (2006). It's a long way to Monte Carlo: Probabilistic display in GPS navigation. *In Proceedings of Mobile Human-Computer Interaction, Mobile HCI'06* (pp. 89-96). Helsinky, Finland.

KEY TERMS

Continuous Control: A continuous control system measures and adjusts the controlled quantity in continuous-time.

Gestural Interfaces: Interfaces where computers use gestures of the human body, typically hand movements, but in some cases other limbs can be used, for example, head gestures.

Haptic Interfaces: Convey a sense of touch via tactile or force-feedback devices.

Manual Control: A branch of control theory that is used to analyse human and system behaviour when operating in a tightly coupled loop.

Nonspeech Sound: Audio feedback, that does not use human speech. The use of nonspeech sound in interaction has benefits such as the increase of information communicated to the user, the reduction of information received through the visual channel, the performance improvement by sharing information across different sensory modalities.

Prediction Horizon: How far ahead the model predicts the future. When the prediction horizon is well matched to the lag between input and output, the user learns how to control the system more rapidly, and achieves better performance.

Quickened Displays: Displays that show the predicted future system state, rather than the current measured, or estimated state.

Sonically Enhanced Interfaces: Interfaces where sound represent actions or content.

Sonification: The use of nonspeech audio to convey information or perceptualize data.

Sound Localisation: The act of using aural cues to identify the location of specific sound sources.

Chapter XXX
Unobtrusive Movement Interaction for Mobile Devices

Panu Korpipää
Finwe Ltd., Finland

Jukka Linjama
Nokia, Finland

Juha Kela
Finwe Ltd., Finland

Tapani Rantakokko
Finwe Ltd., Finland

ABSTRACT

Gesture control of mobile devices is an emerging user interaction modality. Large-scale deployment has been delayed by two main technical challenges: detecting gestures reliably and power consumption. There have also been user-experience-related challenges, such as indicating the start of a gesture, social acceptance, and feedback on the gesture detection status. This chapter evaluates a solution for the main challenges: an event-based movement interaction modality, tapping, that emphasizes minimal user effort in interacting with a mobile device. The technical feasibility of the interaction method is examined with a smartphone equipped with a sensor interaction cover, utilizing an enabling software framework. The reliability of detecting tapping is evaluated by analyzing a dataset collected with the smartphone prototype. Overall, the results suggest that detecting tapping is reliable enough for practical applications in mobile computing when the interaction is performed in a stationary situation.

INTRODUCTION

The source of innovations in a mobile device user interface lies in combinations of input and output technologies that match the user's needs. In the mobile context, movement sensing, and haptic feedback as its counterpart, offers a new dimension to multimodal interactions. There are use cases where traditional interaction modalities are insufficient, for example, when the device is placed in a pocket or a holster, or if the user is wearing gloves. In these situations the user cannot press or see buttons to interact with the device. Instead, small motion gestures can be used as a limited, but convenient, control modality. The movement of the device can be captured with a 3-axis accelerometer, and the resulting acceleration signal can be used to detect the movement patterns for controlling the device.

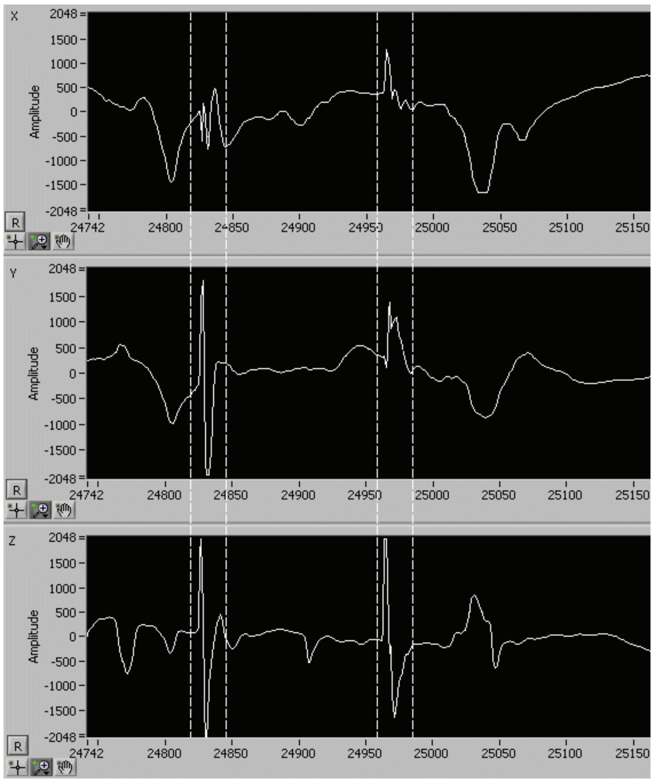
One of the main questions in the application of a movement-based interface is how to distinguish gesture movements the user performs from those movements that are produced by various other activities while carrying and using the device. Reliability can be argued to be the most important challenge in developing a mobile device gesture interface. This chapter presents a reliability evaluation of an unobtrusive event-based gesture interface by analyzing a multiuser dataset collected with a smartphone prototype. Another main challenge has been the relatively high power consumption from the continuous measurement of acceleration, which is not acceptable in mobile devices. Novel accelerometers are capable of producing interrupts based on exceeded thresholds; therefore, the detection, initiated by a hardware interrupt, can be implemented as event based and low power. The technical feasibility of event-based tapping detection is examined with a smartphone equipped with a sensor interaction cover, Figure 1, and an enabling software framework. Furthermore, the chapter addresses the issue of flexibly customizing the gesture interface and feedback modalities relevant to aiding the user.

There are various ways of implementing a gesture interface. This chapter focuses on analyzing the tapping interaction, which shows potential as a significant application of accelerometers in future mobile devices. More specifically, the chapter addresses the movement pattern where the user taps the device twice consecutively, which is called a double tap. With an implementation based on abstractions initiated by sensor-driven interrupts, the aim is a low-power, reliable, and customizable user interaction modality.

Gestures can be detected either from a continuous stream or discrete segments of sensor data. In detection from discrete segments, gesture start and end are explicitly marked with a button instead of a continuous flow of device movements. From

Figure 1. Smartphone prototype equipped with the sensor interaction cover

Figure 2. Three channels of acceleration data on a double tap performed while walking. The Z axis has two distinguishable spikes in this double tap.



the usability perspective, interaction without explicit marking is preferred, in general, since it requires less attention from the user. However, continuous data streaming and execution of the gesture detection algorithm requires continuous data processing, which normally consumes battery power.

The development in digital acceleration sensor technology enables the integration of programmable interrupt-based solutions that can operate with low current consumption. Such sensors generate interrupts when acceleration on a spatial axis is over or below a set threshold level. Hence, movement detection algorithms, initiated by an exceeded threshold, can be implemented as event based instead of continuously processing a stream of data. The processing load at the mobile device side is similarly reduced since the operating system is woken up less frequently. This development opens up new possibilities for practical application of the technology in mass products such as mobile phones.

The distinguishable form of the tapping pattern, processed after the event threshold, is the basis for the potential reliability of detecting them, even when the detection process is continuously active, Figure 2. By contrast, free-form gesture recognition has a much wider problem setting, requiring a more complex model of the gesture and thus, heavier processing load, making continuous processing much more challenging, especially in mobile devices.

This chapter publishes the first statistical performance evaluation based on a dataset that characterizes the reliability of user-independent tapping interaction in mobile phones. Moreover, the sensitivity of the method to misrecognitions is evaluated with scenarios consisting of various activities. As an introductory topic, applying a smartphone equipped with sensor interaction cover, customization, and feedback of the addressed interaction modality are discussed.

BACKGROUND

In acceleration sensor-based gesture recognition, gestures are detected either from a continuous stream or from discrete segments of sensor data. While this chapter addresses the detection of movement patterns from a continuous stream, there are a lot of studies in the literature on gesture recognition from discrete segments (Feldman, Tapia, Sadi, Maes, & Schmandt, 2005; Mäntyjärvi, Kela, Korpipää, & Kallio, 2004). Specifically, acceleration sensors have been applied in user-trainable and pretrained machine-learning-based gesture recognition systems (Kallio, Kela, Korpipää, & Mäntyjärvi, 2006; Kela, Korpipää, Mäntyjärvi, Kallio, Savino, Jozzo, & Di Marca, 2006). Free-form gesture recognition still has a limitation; it requires an explicit marking of the gesture, for example, with a button, and longer duration gestures to increase the recognition accuracy. Hence the interaction requires more user effort, and gesturing can be socially obtrusive. However, despite the possible obtrusiveness when applied in public places, free-form gestures also have a wide range of potential uses in other settings, such as games, home electronics control, and so forth, where social acceptance does not limit the use of the modality. The social aspect, distinctively important in the mobile usage context, has been addressed by Linjama et al. (Linjama, Häkkilä, & Ronkainen, 2005), Rekimoto (2001), and Ronkainen et al. (Ronkainen, Häkkilä, Kaleva, Colley, & Linjama, 2007). Based on the literature, it can be extrapolated that, when performed with a mobile device such as a phone, smaller gestures are considered more socially acceptable than large ones.

This chapter especially advocates the unobtrusiveness of the interaction; gestures as small and as unnoticeable as possible are preferred, assuming they are more acceptable by the users (Linjama et al., 2005). Examples of possibly useful small-scale gestures include shaking the device, for example, Levin and Yarin (1999), and swinging it from side to side (Sawada, Uta, & Hashimoto, 1999). However, both of these interaction methods can be considered quite noticeable, regardless of scale. Shaking also raises the question of how many repetitions of the shake movement are required until a shake is recognized. A simple accelerometer-based tilting control has been discussed in the literature in many studies over the years, for example, Rekimoto (1996), but also recently, for example, combining tilt and vibrotactile feedback (Oakley, Ängeslevä, Hughes, & O'Modhrain, 2004), scrolling, and switching between landscape and portrait display orientations (Hinckley, Pierce, Horvitz, & Sinclair, 2005). Tilting is another potentially unobtrusive, and very simple to implement, movement-based interaction modality to be applied in carefully selected use cases in mobile computing.

A minimalist extreme in hand gestures is tapping the mobile device, first introduced in Linjama and Kaaresoja (2004). Tapping only requires a small scale of device movement, and can be performed by finger or palm. The technological benefit is that tapping can be relatively straightforwardly captured with a 3-D accelerometer, since the resulting movement pattern has a distinguishable sharp spike form. The detection problem can be narrowed down by applying a small, predefined fixed set of movement patterns: tap events.

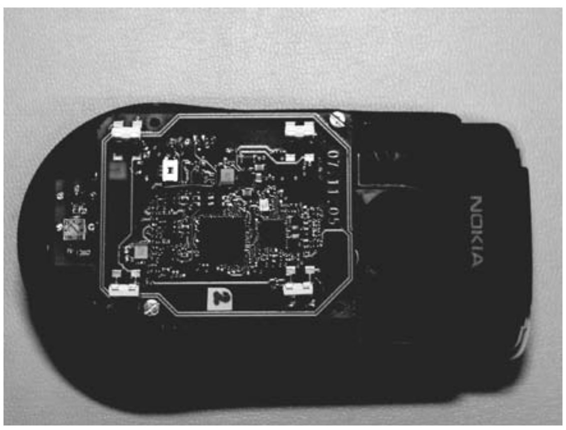
The unique usability benefit of the tap interaction is that it is discreet and can be used if the mobile device is located in a pocket or a backpack, since explicit marking is not needed. Furthermore, the user is not required to hold the device or see the keyboard to interact. A good example of a use case where tapping is useful can be found in the Nokia 5500 phone (Nokia, 2006): when a text message arrives, the user has 30 seconds to tap the phone twice and the message will be read aloud to the user. It is useful when the phone is in a pocket or on a belt, or the user is wearing gloves; the message can be read without first taking the phone into the hand and opening the keypad lock. Furthermore, tapping can be used as an additional modality. For instance, phone music player commands, such as play next or previous song, can be controlled by tapping on either side of the phone, which is convenient when the device is worn on a belt or in a pocket. Again, the user does not have to take the phone, open keypad lock, and press a button to perform the control action.

SENSOR INTERACTION COVER

Interrupt-initiated abstracting of movement patterns can be performed using a separate microcontroller, or, ideally, it can be directly integrated in the sensor chip. A sensor interaction test platform was developed to experiment with the interaction concept. The platform consists of a Symbian S60 phone (Nokia 6630) equipped with a sensor and feedback cover attached to the back of the smartphone, Figure 3.

Inside the cover, the hardware includes a 3-D acceleration sensor (STMicroelectronics LIS3LV02DL), a microcontroller (Atmel), an NFC reader, blue LEDs, a buzzer, and a vibra motor, Figure 2. The board is two sided. Tap detection parameters and feedback configuration can be set to the microcontroller from the phone software. The tap detection algorithm and the feedback processing are performed in the cover microcontroller, and the cover transmits recognized tap events to the phone through USB. Thus, the communication between the cover and the phone, as well as power consumption, is minimized.

Figure 3. Sensor interaction cover hardware



INTERACTION CUSTOMIZATION

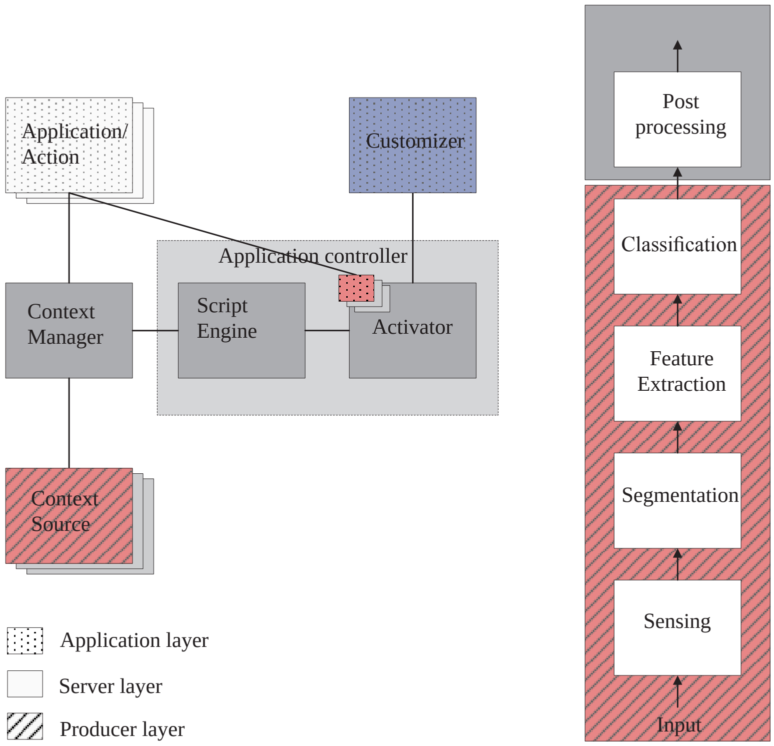
Once sensor events are abstracted by the microcontroller and sent to the phone through the USB, they should activate the desired actions in the mobile phone. Flexibly connecting the abstracted sensor events to various application actions requires supporting middleware on the phone side. Instead of connecting an application directly to a device driver, the data is abstracted into a uniform representation applied through Context Framework.

Context Framework (CF) is a blackboard-based software framework for enabling and customizing situation-aware and sensor-based mobile applications (Korpipää 2005; Korpipää, Mäntyjärvi, Kela, Keränen, & Malm 2003). All interaction-related information, including implicit and explicit sensor-based inputs, is treated as context objects within the framework, expressed with a uniform vocabulary. An implemented instantiation of the framework is illustrated in Figure 4 (left-hand side). In this case, the sensor signal abstracting process functionality is on the microcontroller side, illustrated in Figure 4 (right-hand side flow diagram). Sensing, feature extraction, and classification are performed at the cover's microcontroller. Classified movement (context) events are sent over the USB to the phone side, where CF enables controlling any available application action based on the events.

The user can create desired context-action behavior with a mobile phone by creating XML-based rule scripts with the graphical UI of the Customizer. CF handles the background monitoring of context events and the triggering of actions according to the rules. The Application Controller facilitates the application control inference on behalf of the user or application, in other words, provides an inversion of control. The framework completely separates the management of sensor-based context events from application code and the hardware. Hence, by applying CF, no changes need to be made to existing mobile phone applications when they are augmented with sensor-based features.

In the case of tapping input, the events are abstracted into context objects by the sensor cover

Figure 4. Context Framework (CF) architecture example instantiation (left), and pattern recognition flow (right)



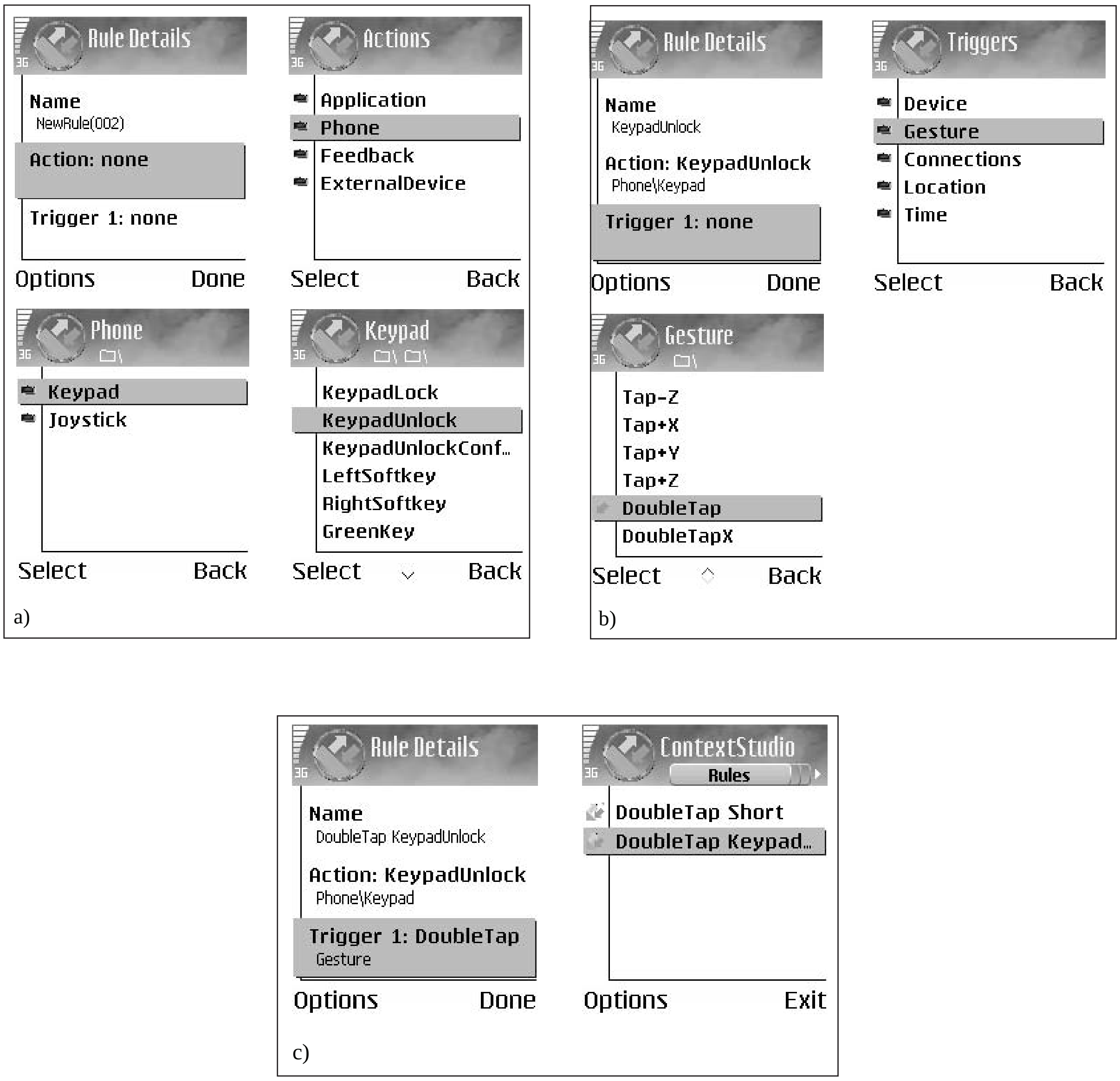
of the phone and delivered to CF. The application developer or the user interface designer can use the Customizer tool to define which application actions are executed by which abstracted sensor events. The definable actions include available feedback modalities, such as tactile, auditory, and visual indications. By creating rules with the Customizer tool, the user can define actions on an operating system level, or for a specific application, by setting a condition part of a rule to include a specific foreground application. For instance, the following accelerometer-based features were defined and executed simply as XML-based rule scripts:

- Playing the next or previous song in music player using double tap
- Activating display illumination using tap
- Unlocking the keypad using double tap

Figure 5 presents a series of screenshots from the Customizer tool, illustrating the definition of a rule that enables the user to unlock the keypad by double tapping the phone.

In Figure 5a, the user selects an action for the rule by navigating through the action type Phone. Keypad and selecting the action value KeypadUnlock. In Figure 5b, the user selects a trigger for the rule by navigating through context type Gesture and selecting context value DoubleTap. The first screenshot in Figure 5c shows the complete rule after the user has selected the elements. After the user selects the option Done, the rule script is generated, and the rule is activated and functional in the context framework. The second screenshot in Figure 5c shows the main rule view with the list of active rules. When the rule conditions are met, the Context Framework automatically performs the action.

Figure 5. Series of screenshots illustrating how to program the phone to open the keypad lock with a double tap.



USE CASES AND USABILITY

Evaluating the general usability is an essential aspect in developing tapping interaction, in addition to evaluating the reliability and technical feasibility. As an extensive topic, however, it cannot fit within the scope of this chapter. The purpose of this section is to briefly discuss a few usability-related points as an overview of the experiments studying the usability of the novel interaction modality.

New interaction modalities, like tapping, have certain application areas where they add value, both in terms of utility (usability) and joy (fun of use). The only way of evaluating these aspects is to try the interaction elements in practice, with real hardware and applications. The smartphone sensor interaction cover serves this purpose. It enables the rapid prototyping and iterative development of interaction concepts and demonstrations. User experiences and feedback can be collected during various stages of development, early concept tests, peer evaluations, and end user tests in the lab and in the field.

A number of formal user tests, to be published separately, have been conducted with the smartphone prototype. The tests measure the potential usefulness of tap interaction with a standard Symbian Series 60 phone user interface. For instance, controlling a phone music player with tapping commands, in addition to existing keypad controls, has been studied. The results from the studies indicate that it is very important to maintain consistency in case there are multiple different uses of tap interaction in several applications of the same device. Users may get confused if tapping is used for too many different purposes, such as muting the phone audio in one application and selecting the next or previous music track in another application. Furthermore, using different tap directions, for example, tapping on the device top or the side for activating different controls in different applications, requires delivering specific instructions to the users.

User satisfaction, joy of use, has also been addressed in the usability tests. The enjoyability of the user experience is largely determined by the very details of the interaction; what kind of feedback elements support the user interaction. What is the metaphor behind the observed device behavior that the user learns when using new interaction modalities? All sense modalities in multimodal interaction must be addressed together.

CONTINUOUS DETECTION RELIABILITY: EXPERIMENTS

This study focuses on analyzing the reliability of detecting double taps in various usage situations. The experiments to be presented next aim to answer how accurately double taps can be detected in a general mobile usage setting and how many misrecognitions occur. The results should reflect an essential part of how feasible this interaction method could become, from the reliability viewpoint, when used in mobile phone applications. Detection accuracy is quantitatively analyzed based on acceleration data collected from users performing the interaction, and the results are discussed.

There were 11 users performing the interaction and the scenarios; 7 of the users were male and 4 were female, aged from 25 to 36 years. The subjects were selected randomly from acquaintances of the authors. The subjects were not interviewed and no subjective opinions were collected, only acceleration signals. Therefore, the limited variability of the subjects in the user group was assumed not to bias the results significantly.

There are two categories of use cases for continuous detection of movement events. In the first category, the detection process is initiated by a specific application or a situation, and is active for a certain time. In the other category, the detection process is always active. In the first category, the use cases can be designed such that misrecognitions, false positives, have a minimal effect. In the latter category, false positives usually have a more negative effect since they may result in incorrect operation. In both cases, the sensitivity of detection should yield enough correct recognitions, true positives, to be acceptable for the users.

This section describes the experiments aiming at evaluating how well the tapping interface performs from a statistical point of view, based on collected data. Detecting tapping events is a type of pattern recognition problem (Duda, Hart, & Stork, 2001), although not a very complex one. The aim of the data analysis during the development process was to reach optimal recognition of a double tap pattern, that is, to find detection algorithm parameters that produce a minimal number of false positives while maintaining a high percentage of true positives. The primary goal was to minimize misrecognitions. The algorithm should give the best results as an average when performed by multiple users, not just one specific user. In other words, the aim is to reach optimal user-independent detection accuracy. The experiment involved collecting a dataset on the target patterns performed by several users in controlled stationary conditions. Furthermore, data from several real-world scenarios containing various daily activities was collected to find out how often misrecognitions occurred.

Data Collecting

In order to evaluate the tapping detection reliability statistically, a sufficiently large dataset is required. Dataset size and the variation it contains are in direct relation to the evidence to support generalization. Data was collected in three stages with the sensor cover-equipped smartphone prototypes. The first stage involved exploring a wide set of activities by a user carrying 1-2 prototypes to find out whether there were any specific activities that produced a lot of false positives. The dataset was collected by one user, and the total duration of the activities in the dataset was 5 hours 8 minutes.

The second stage involved having several users perform the target patterns in stationary controlled conditions involving no other activities. This dataset consisted of 11 users performing double-tap patterns. Data was collected in three categories, arranged by the user's skill level and the given advice. The user groups were beginner, people who had never heard about tapping, and advanced, people who knew or were informed about how the tapping user interaction works. There were six users in the beginner group. In the beginner group, the users were only given one piece of advice: to perform the tapping with their hand(s), not by tapping the phone on the table. The second group, five advanced users, were first told to use one hand for tapping and next to use both hands, that is, hold the phone with one hand and tap with the other. Figure 6 shows an example of both ways of tapping interaction.

There were five users, the same ones, in both of the advanced groups. In the three categories, each user performed a double tap 18 times, resulting in total target of 288 repetitions in the dataset. Repetitions were performed in phases of three repetitions and a break, during which the device was put on the table to avoid a routine speed-up and fixation on a certain way of interaction.

The third stage involved having several users perform scenarios involving real-world activities while carrying the prototype in their pocket. The purpose of this dataset was to find the occurrence of false positives during the scenarios, on average over multiple users. There were four to five users in each of the scenarios. The total length of the activity dataset was approximately 54 minutes. The tapping pattern has a sharp spike-form shape, and proper detection requires a relatively high sampling rate. Hence, the total amount of raw data collected for this experiment was approximately 68 megabytes.

Figure 6. Tapping performed with one hand (a) and with both hands (b)

(a) *(b)*

EXPERIMENT RESULTS

The collected acceleration data was used to analyze the tapping interaction from multiple aspects. The experiments focused on a specific form of tapping, a double tap. Double tap means performing two consecutive taps within a certain short time span, much like a double click with a mouse. Each aspect of this interaction studied with the collected dataset is described in detail in this section. The experiments produced numerical measurements of the system's tapping detection accuracy. The measurements are briefly introduced here before presenting the results and analyzing them.

The first experiment was an initial pilot test, which was designed to count the number of double-tap patterns detected where they should not exist. In other words, the experiment measures the occurrence of false positives, which can be reported as a number per time unit. For example, the aim could be that there is no more than one false double-tap detection per hour.

False positives can also be represented in relation to how many patterns could be falsely detected from a dataset. The relative number of false positives in a dataset can be given by dividing the occurrence of all detected false positive patterns with all segments of data where there should not be a detected pattern. Here a segment is defined as the maximum time span required to detect one pattern. For example, for double tap pattern the maximum allowed duration is 1.1 seconds. This is due to the algorithm wait time for the second tap to appear after the first one. For instance, in a dataset of 110 seconds, there are 100 segments that could potentially contain a double tap. One false double tap in that dataset would result in one percentage of false positives.

True positive means a correctly detected pattern, for example, a double tap is detected correctly when it is performed by the user. The relative occurrence of true positives can be given by dividing all detected true positive patterns by all actually performed true patterns in a dataset.

Pilot Test

The goal of the pilot test was to explore whether some of the randomly selected ordinary daily activities would produce a high occurrence of false positives. This experiment did not contain any actual double taps performed by the user. The user was assigned to carry one or two prototypes in a pocket during various daily activities, for example, random outdoor activities (cleaning the yard, commuting, driving a car, walking, jogging, biking, cross-country skiing, and roller-skating). The users were free to select which clothes to wear and which pockets to carry the devices in. The tasks were given as, for example, "take the phone with you and go jogging." Table 1 summarizes the results of this test.

There were several activities that did not produce any false positives, such as jogging, various outdoor activities, biking, going for lunch, and roller-skating. The activity that produced the most false positives was cross-country skiing.

Overall, the test indicated that potential problem areas are accidental tapping by hand, ski stick, backpack, and so forth, and when the phone is laying freely on a moving and trembling surface such as a car dashboard. After the pilot test, the detection algorithm and parameters were adjusted to reduce the misrecognitions.

Stationary Conditions

Next, an experiment was performed in controlled stationary conditions. The purpose of the experiment was firstly to gain validation of how well the target patterns are recognized in a stationary situation when there are no external disturbances. Secondly, it is important to know whether there are differences between two groups of users when one has no idea what a tapping interface is and the other has prior knowledge of how to interact with tapping. The results indicate different variations in the first-time use of tapping in terms of gesture signal waveform and the detection accuracy. Thirdly, the results show whether there are major differences between individual users, whether

Table 1. Occurrences of double-tap false positives during random daily activities

Activity	**Phone numbers, placement**	**Duration (min)**	**False positives**
Commuting (dressing, driving, walking, stairs up, stairs down, office)	2, left and right lower jacket pocket	28	2
Travel by car, tarmac road	1, dashboard	70	1
Travel by car, rough gravel road	1, dashboard	20	2
Jogging	2, left and right jacket chest pocket	5	0
Cross-country skiing (walking, changing, skiing, walking, undressing)	2, jacket pocket, backpack	75	7
Outdoor activities (removing snow, walking, putting bike in storage)	2, left and right jacket chest pocket	5	0
Biking (gravel road and tarmac road)	2, jacket pockets	10	0
Going for lunch (stairs down, lunch, walking, stairs up)	1, jeans pocket	25	0
Roller skating	1, loose short trousers front pocket	35	0
Roller skating with sticks	1, pants front pocket	35	1
Total		**5 hours 8 min**	**13**

Table 2. Recognition rate in stationary situation for various user groups in 3-axis detection

User group	**Users**	**True positive %**
Beginner	6	55.2
Advanced one hand	5	90.6
Advanced both hands	5	90.2

the interaction is equally assimilated by all users or if there are some individuals that cannot use the method as well. Finally, the interaction by tapping can be performed by using one hand or both hands, and the results indicate which is preferred from the reliability point of view with the evaluated algorithm.

The results can be calculated in two ways: the interaction can be allowed from any of the three axes, or from one selected axis only. In most single application use cases, the direction of tapping is known in advance and can be restricted. For example, music player next and previous commands can only be initiated with a tap on either side of the phone, by utilizing only the x-axis while disregarding the others. Hence, depending on the use case, it is feasible to filter the data from one or two other axes and apply the signal from one axis only. The results are first presented for 3-axis detection, Table 2.

The results show that double taps can be detected fairly well in stationary conditions, except

Table 3. Recognition rate in a stationary situation for each individual user in 3-axis detection

User group	User1	User2	User3	User4	User5	User6	Total
Beginner	94.4	0	77.8	88.9	0	73.3	**55.2**

User group	User1	User2	User3	User4	User5	Total
Advanced one hand	88.2	73.3	100	88.2	100	**90.6**
Advanced both hands	100	50.0	100	100	100	**90.2**
Total	**94.1**	**61.7**	**100**	**94.3**	**100**	**90.4**

in the beginner group. The difference between the beginner and advanced user groups is quite large, which suggests that first-time users may have trouble when starting to apply the method if they are not properly informed. There were also distinct differences between the individual beginner users, Table 3.

The data from the beginner users that produced low accuracies revealed that they performed the taps too lightly. Half of the beginner users chose to perform the tapping with one hand, and half with both hands. One beginner user tapped the top of the device and one the bottom, others from the side. The two beginner users that tapped with one hand had the zero results. The recollection from the actual test situation and data visualization confirm that the two one-hand users having a zero result only touched the device very lightly instead of properly tapping it. In other words, the first-time users' low performance is partly an algorithm sensitivity issue, but most importantly it is due to the lack of information the user has on how to do the tapping in the first place. The results can be improved by modifying the parameters to be more sensitive, but then the false positives tend to increase. The most straightforward way to improve the result is simply to advise the beginner users to tap with the correct intensity. Feedback is one way of giving immediate information to the user.

It must be noted that this experiment produced no information on the learning curve; it simply provides data on how differently first-time users may perform the gesture. There was no feedback or interaction in the test to guide the user on how to improve. In this sense it was a "blind" blank test to examine different users' approaches to performing a double tap, as interpreted from the signal waveform and the resulting detection accuracy. In a normal usage situation, the user would learn that too light taps do not cause the desired operation, and would likely either modify their behavior or abandon the method. In this test the users did not know that they tapped too lightly and thus, could not know how to change their tapping style.

The results in Tables 2 and 3 present the results for a setting where double taps from any direction are allowed. Table 4 presents results where only one predetermined axis signal is applied to detect a double tap. A significant increase in detection accuracy is evident. Furthermore, it is likely to reduce the occurrence of false positives, although it was not tested in this study. In light of

Table 4. Advanced user recognition rate in a stationary situation in 1-axis detection

User group	Users	True positive %
Advanced one hand	5	95.3
Advanced both hands	5	98.8
Total	**5**	**97.0**

the results, it is preferable to restrict the detection axis whenever is possible.

The results indicate that tapping is detected slightly more accurately when performed with one hand in 3-axis detection. In 1-axis detection, the accuracy is slightly better when performed with both hands. However, statistically, a conclusion cannot yet be drawn with this dataset on which way of tapping is more reliable.

False Positives - Multiple Users

The purpose of the experiment with mobile scenarios was to find the occurrence of false positives during the selected common daily activities: walking, walking up stairs, jogging, and roller-skating. Furthermore, the scenarios were performed by multiple users in order to address the issue independent of the user. The scenarios in the experiment were designed to address a usage situation where the phone is in the user's pocket and the user could tap the phone from any direction. The users wore their own clothes and were free to select where to put the phone during the test. No other hard objects were allowed in the same pocket.

The results show that the number of double tap false positives was zero during the total of 54 minutes of activity data. By adjusting the algorithm parameters to more sensitive (which also increased true positives in the stationary test), false positives started to occur. The most false positives occurred on stairs. However, the parameter set that produced zero misrecognitions was generally perceived as sensitive enough, even though there were beginner users who would have benefited from increased sensitivity.

Summary of Results

Overall, the results based on the collected data, Table 5, indicate that detection is reliable enough for practical applications in mobile computing when the user performs the interaction in a stationary situation. Moreover, the number of false positives is low enough for types of mobile applications with at least a restricted scope. The

Table 5. Overview of the test results

Test	Users	True positive %
Beginner	6	55.2
Advanced one hand 3 axis	5	90.6
Advanced both hands 3 axis	5	90.2
Advanced one hand 1 axis	5	95.3
Advanced both hands 1 axis	5	98.8

results have significance for commercial applications built on use cases that have a clear usability advantage from the tapping interaction.

The results also show that there is room for improvement. This especially concerns the usability aspect of first-time use. An important question is how to give instruction on using the interface. This experiment took a worst-case scenario where the user was given almost no information, much like when the user does not even read the manual before starting to use the device. In a real learning situation, however, the user may sometimes even look for instructions in the manual, or someone will demonstrate how to use the feature. Thus, the results could be different. Furthermore, unlike in this test, the user would get feedback if the device did not respond to the interaction. Analyzing the learning curve, which is another relevant topic, requires a different experiment setup.

Having zero misrecognitions from four activities performed by four to five users with a total of 54 minutes of data does not yet statistically allow a strong generalization statement, although it is a good result, and shows that practical application is certainly feasible. To gain even wider evaluation, the next phase is to perform longer tests by equipping the users with prototypes for use in their normal daily lives.

FUTURE TRENDS

Although this study did not specifically discuss the user experience side of movement-based interaction, there is one aspect we would like to briefly address when viewing future trends: feedback.

This aspect is still often found insufficient in novel user interfaces. While the presented experiments evaluated the reliability of the double-tap detection, future work includes analyzing the learning curve, the best type of feedback, and its effect on the user experience.

The user experience and learning curve for new interaction modalities can potentially be improved with suitable feedback. For example, if the beginner user makes too light taps in a tutorial mode, the device can indicate this with feedback. In general, feedback gives an indication of the state of the system and guides the users in how to use it. As suggested by O'Modhrain (2004), a key to the design of successful touch and haptic-based mobile applications is in ensuring a good mapping between the tasks, the required sensory cues, and the capabilities of the system on which the application is to be implemented. With the Customizer tool, introduced earlier, developers and user interface designers can easily experiment with different multimodal input and output combinations to find the most suitable and enjoyable solution for their application needs.

Different combinations of the feedback patterns (vibration, LED, sound) available in the interaction cover were implemented in this study. The option of using direct cover feedback in addition to phone vibra in the interaction had the benefit of avoiding possible latencies in feedback generation on the phone side. The vibra feedback was thus precisely adjustable to the desired parameters. Even though experiments on feedback supporting usability were not presented in this chapter, it can be predicted that utilizing minimalist gesture control, together with related haptic feedback elements, has great potential in a mobile device usage and technology context. Haptic content fidelity can be rather low if it is designed to be multimodal; visual and haptic content are applied synchronously to support each other. The interaction and content design are used to promote the adoption of the technology among users.

Continuous detection of small sharp movement events also facilitates forms of gestures other than double tap. As an analogy to mouse control, there is a click and a double click. Obviously, single taps can be utilized for many purposes. However, single taps are more sensitive to various disturbances, such as accidental knocking, dropping, quick swings, turning, and so forth, that can produce a similar sharp pattern to the data and thus, a false positive. Another interesting gesture that feels natural is to swing the device. There are many other possible movement patterns to utilize in the future.

Several research questions remain, such as how to inform the user about the correct intensity of the tapping, and what kind of learning curve the tapping has. Many of the misrecognitions in the beginner group, as well as in the group that used only one hand, were due to too light a touch when tapping the device. In the beginner group, the gestures were even confused with touching in a user's approach. From the detection algorithm point of view, there is a trade-off: the parameters cannot be set too sensitively to avoid increasing the occurrence of false positives. Even though a lighter tap is viewed as more satisfying by some users, this usability increase cannot cost the reliability too much.

Yet another relevant research problem is to examine the recognition accuracy of target patterns during various activities in mobile usage. This study addressed the stationary situation and false positives during scenarios. A relevant question is what happens if the user performs the interaction while jogging, for example, without stopping to do it. Future work includes examining whether and how the continuous gesture interaction algorithms should adapt to the movement situation of the device.

CONCLUSION

Gesture control is increasingly being applied in mobile interaction. Widespread movement interaction application in mobile devices has been delayed by research challenges such as reliably detecting gestures, power consumption, and user experience-related issues such as obtrusiveness and increased effort. This chapter has focused on analyzing and evaluating the reliability of an event-based gesture interaction modality that emphasizes minimal user effort in interacting with a mobile device. The technical feasibility of the interaction modality was examined with an implementation in a smartphone environment. The reliability of continuous detection of sharp movement events produced by the user by lightly tapping the phone was evaluated by analyzing a dataset collected with the prototype.

The results show that for five informed users performing 36 repetitions of double taps in controlled stationary conditions, the target pattern was recognized with 90.4% accuracy for 3-axis detection and 97.0% for 1-axis detection. In four mobile scenarios containing 54 minutes of daily activities, each performed by four to five users carrying the prototype, there were no false positive detections of the pattern. Overall, the results based on a statistical analysis of the collected acceleration data suggested that double-tap detection is reliable enough for practical applications in mobile computing when the user performs the interaction in a stationary situation. Furthermore, it was found that the occurrence of false positives is low enough for application, presuming carefully selected usage situations where possible misrecognitions are not critical. The contribution of this work has significance for commercial utilization.

Several research questions remain to be addressed as future work. These include how to inform the user about the correct intensity of tapping; there were users with too light a touch in the experiments. From the detection algorithm point of view, a balance needs to be found as the parameters cannot be set too sensitive to avoid increasing the occurrence of false positives. Another important research problem is to examine the recognition accuracy of target patterns during various activities in mobile usage. This study addressed the stationary situation and false positives during scenarios.

As to the movement interaction detection performance in general, the trend of development firmly aims toward increased reliability. As a result, the restricted application-specific use cases are likely to be followed by more general platform-level operations, where movement can be used as an additional interaction modality complementary to the existing ones. With emerging commercial utilization, it is easy to see the beginnings of wider adoption of the new interaction modality in mobile computing, while not forgetting that there is still further work to be done.

ACKNOWLEDGMENT

We would like to acknowledge the work of Arto Ylisaukko-oja in the hardware development, Hannu Vasama for designing the cover casing, and other contributors at Finwe, Nokia, and VTT for their kind collaboration.

REFERENCES

Duda, R., Hart, P., & Stork, D. (2001). *Pattern classification* (2nd ed.). John Wiley & Sons.

Feldman, A., Tapia, E. M., Sadi, S., Maes, P., & Schmandt, C. (2005). ReachMedia: On-the-move interaction with everyday objects. In *Proceedings of IEEE International Symposium on Wearable Computers (ISWC'05)* pp. 52-59.

Hinckley, K., Pierce, J., Horvitz, E. & Sinclair, M. (2005). Foreground and background interaction with sensor-enhanced mobile devices. *ACM Transactions on Computer-Human Interaction (TOCHI), 12*(1), 31-52.

Kallio, S., Kela, J., Korpipää, P., & Mäntyjärvi, J. (2006). User independent gesture interaction for small handheld devices. *Special Issue on Intelligent Mobile and Embedded Systems of IJPRAI, 20*(4), 505-524.

Kela, J., Korpipää, P., Mäntyjärvi, J., Kallio, S., Savino, G., Jozzo, L., & Di Marca, S. (2006). Accelerometer based gesture control for a design environment. *Personal and Ubiquitous Computing* (pp. 1-15). Online First Springer.

Korpipää, P. (2005). *Blackboard-based software framework and tool for mobile device context awareness.* Ph.D dissertation. VTT Publications

579. Retrieved from http://www.vtt.fi/inf/pdf/publications/2005/P579.pdf

Korpipää, P., Mäntyjärvi, J., Kela, J., Keränen, H., & Malm E-J. (2003). Managing context information in mobile devices. *IEEE Pervasive Computing Magazine*, *2*(3), 42–51.

Levin, G. & Yarin, P. (1999). Bringing sketching tools to keychain computers with an acceleration-based interface. In *Proceedings of the CHI 98* (pp. 268-269). ACM: New York.

Linjama, J., Häkkilä, J., & Ronkainen, S. (2005, April 3-4). Gesture interfaces for mobile devices—Minimalist approach for haptic interaction. Position paper in *CHI 2005 Workshop "Hands on Haptics."* Portland, Oregon. Retrieved from http://www.dcs.gla.ac.uk/haptic/sub.html

Linjama, J., & Kaaresoja, T. (2004). Novel, minimalist haptic gesture interaction for mobile devices. In *Proceedings of the NordiCHI2004* (pp. 457-458). ACM Press.

Mäntyjärvi, J., Kela, J., Korpipää, P., & Kallio, S. (2004). Enabling fast and effortless customization in accelerometer based gesture interaction. In *Proceedings of the International Conference on Mobile and Ubiquitous Multimedia (MUM)* (pp. 25–31). ACM.

Nokia Corporation. 5500 phone. (2006). Retrieved from http://europe.nokia.com/link?cid=EDITORIAL_8657

Oakley, I., Ängeslevä, J., Hughes, S., & O'Modhrain, S. (2004). Tilt and feel: Scrolling with Vibrotactile Display. In *Proceedings of Eurohaptics* (pp. 316-323).

O'Modhrain, S. (2004). Touch and go - Designing haptic feedback for a hand-held mobile device. *BT Technology Journal*, *22*(4), 139-145.

Rekimoto, J. (1996). Tilting operations for small screen interfaces. In *Proceedings of the 9th Annual ACM Symposium on User Interface Software and Technology* (pp. 167-168).

Rekimoto, J. (2001). GestureWrist and GesturePad: Unobtrusive wearable interaction devices, In *Proceedings of the Fifth International Symposium on Wearable Computers (ISWC)* (pp. 21-27).

Ronkainen, S., Häkkilä, J., Kaleva, S., Colley, A., & Linjama, J. (2007). Tap input as an embedded interaction method for mobile devices. In Proceedings of the First Tangible and Embedded Interaction (pp. 263-270). ACM: New York.

Sawada, H., Uta, S., & Hashimoto, S. (1999). Gesture recognition for human-friendly interface in designer - consumer cooperate design system. In *Proceedings IEEE International Workshop on Robot and Human Interaction* (pp. 400-405). Pisa, Italy.

KEY TERMS

Accelerometer: 3-D accelerometer is a sensor capable of measuring object acceleration along three spatial axes.

Double Tap: Double tap is a form of movement interaction where the user performs two consecutive taps on a mobile device with a finger or palm, each producing a sharp spike waveform in an accelerometer signal measured with a high sampling rate.

Gesture Interaction: Gesture interaction here refers to explicit movements made with a mobile device while holding it in a hand in order to perform any tasks with the device.

False Positive %: False positive percentage is the relative number of falsely detected patterns, given by dividing the occurrence of all detected false positive patterns by all segments of data where a detected pattern in a dataset should not exist.

Pattern Recognition: Pattern recognition is the scientific discipline whose goal is the classification of objects into a number of categories or classes. Objects can be, for example, signal waveforms or any type of measurement that needs to be classified. These objects are here referred to using the generic term "patterns."

Smartphone: A smartphone is an advanced multifunctional mobile phone with a platform open to third-party software.

True Positive %: True positive percentage is the relative number of correctly detected patterns, given by dividing all detected true positive patterns by all actually performed true patterns in a dataset.

Chapter XXXI
EMG for Subtle, Intimate Interfaces

Enrico Costanza
Ecole Polytechnique Fédérale de Lausanne, Switzerland

Samuel A. Inverso
The Australian National University, Australia

Rebecca Allen
UCLA Design | Media Arts, USA

Pattie Maes
MIT Media Lab, USA

ABSTRACT

Mobile interfaces should be designed to enable subtle, discreet, and unobtrusive interaction. Biosignals and, in particular, the electromyographic (EMG) signal, can provide a subtle input modality for mobile interfaces. The EMG signal is generated by a muscle contraction and can be used for volitional control; its greatest potential for mobile interfaces is its ability to sense muscle activity not related to movement. An EMG-based wearable input device, the Intimate Communication Armband, is presented in this chapter to demonstrate this subtle interaction concept. The device detects subtle, motionless gestures from the upper arm. Experimental results show that the gestures are reliably recognized without user or machine training, that the system can be used effectively to control a multimodal interface, and that it is very difficult for observers to guess when a trained user is performing subtle gestures, confirming the subtlety of the proposed interaction.

INTRODUCTION

Mobile communication devices provide ubiquitous connectivity, allowing people to engage in private and personal communication from virtually anywhere. They are often used in public places (offices, libraries, museums, theatres, restaurants) or on public transportation (such as buses and trains), where the user is surrounded by others not involved in the interaction. Using a mobile device in a social context should not cause embarrassment and disruption to the people in the immediate environment. This problem has been reported by social scientists (Fortunati, 2002, Okabe & Ito, 2005), and it is emphasised by the many signs that can be found in public places inviting or ordering people to turn off cell phones. Deactivating these devices is an extreme solution, as it completely annihilates the devices' functions and advantages, and indeed, users are not inclined to do so. Ring-tones' replacement with vibrating alerts in mobile phones constitutes an example of a widespread subtle interface to improve social acceptance, while still allowing access to the device's functionality. Unfortunately, this idea of subtlety and social acceptance has not yet been generalized and is lacking in other parts of the interface design.

Mobile interfaces should be designed to enable subtle, discreet, and unobtrusive interaction. The human-computer interaction (HCI) research community has recently shown increasing interest in the design of mobile and wearable interfaces that are socially acceptable, and that take into account the social context of users. Rekimoto (2001) advocates that, to be accepted in everyday and public situations, wearable input devices should be "as natural and (conceptually) unnoticeable as possible." Lumsden and Brewster (2003) question the social acceptance of speech-based and gesture-based interaction. Marti and Schmandt (2005) address the disruption caused by mobile phone notifications with a subtle notification and vetoing system. The work presented in this chapter extends this research thread, demonstrating how biosignals, and, in particular, the electromyographic (EMG) signal, a biosignal generated by muscular activity, can be used to enable natural and unnoticeable interaction.

This chapter proposes intimate interfaces: discrete interfaces that make interaction with mobile devices private and concealed as much as possible, in order to minimize the disruption of colocated individuals, and let mobile technology gain social acceptance. Even though it has been suggested that making mobile interaction public and evident could help colocated individuals to understand and accept the behaviour of mobile technology users (Hansson, Ljungstrand, & Redström, 2001), users themselves can inform others of their interaction, if they want. What is private can be made public, but not vice-versa.

In a mobile context, users are often involved in a primary activity, ranging from navigation and monitoring of the immediate environment (e.g., waiting for an incoming train) to specific tasks, such as equipment maintenance or field work. Mobile devices are used either for assistance to the primary task–providing access to equipment documentation or data logging–or for *side involvements*, in the sense of collateral activities unrelated but not conflicting with the primary task, as defined by Goffman (1963). Interaction techniques based on EMG signals can provide an extra modality for interaction, one that does not conflict with the primary task.

The EMG signal is generated by muscle contractions and can be used for volitional control. EMG's greatest potential for mobile interfaces is its ability to sense muscle activity not related to movement, allowing the definition of a class of *subtle* or *motionless gestures.*

This chapter covers the design and evaluation of the *Intimate Communication Armband*: a wearable device that detects subtle motionless gestures through EMG signals, and can be used to control existing devices through a Bluetooth interface. The next section provides background about EMG and its applications within human-computer interaction (HCI). Subsequently, the concept of *motionless gestures* and a system to recognize them, the Intimate Communication Armband, are introduced, followed by evaluation through three user studies. The studies assess the basic functionality of subtle gesture recognition, the use of such gestures to control a multimodal interface, and how noticeable the gestures are to bystanders. Finally, suggestions about further work and concluding remarks are presented.

BACKGROUND

The EMG signal is an electrical signal generated by a muscle contraction. Through electromyography, it is possible to sense muscular activity related to movement, such as lifting or folding an arm, and also *isometric* activity: muscular activity that does not produce movement (Tanaka & Knapp, 2002). An example of isometric activity is pushing against a wall; where muscles are activated, but the wall prevents movement; similarly, isometric activity can be produced by flexing the muscles without load, as when "showing off muscles." The sensing of isometric activity has great potential for mobile interfaces, as detailed in the following section.

In the last three decades, biomedical engineering has yielded many effective methods for recording and computer-aided analysis of EMG signals (DeLuca, 1979). This chapter will only consider recording through noninvasive surface electrodes: conductive elements placed on the skin and kept in place, either through adhesive (similar to that commonly found in bandages) or other means (e.g., elastic fabric bands). EMG signals can also be recorded using needle electrodes, introduced through the skin, which produce better signals because they are in close contact with the muscle. While their use can be justified in a medical context, the discomfort that they cause to the user makes them highly impractical for the kind of everyday applications considered in this chapter. Moreover, current integrated circuit technology makes it possible to produce EMG signals from surface electrodes of higher quality than in the past.

Electromyographic (EMG) Signal

The electromyographic (EMG) signal is the result of the superposition of electric voltage generated by each motor unit in a muscle. Being a voltage signal, it is sensed through pairs of differential electrodes, generally located over the muscle of interest, each pair constituting a *channel*. Because surface electrodes record from a large number of motor units at the same time, the resulting EMG signal can be represented as a signal with Gaussian distributed amplitude, typically ranging from 100 μV to about 1 mV (DeLuca, 1979).

Electrodes, Recording, and Applications

Commercial surface electrodes are generally Ag/AgCl plates covered with conductive gel (often solid gel for increased comfort) and attached to the skin with adhesive. The gel is used to improve the electrode to skin interface, lowering the impedance seen from the sensor, and reducing motion artefacts[1]. *Activ*e or *drive*n electrodes are sometimes used to create a feedback control loop between the sensor and the body (Webster, 1992), this method also reduces motion artefacts, eliminating the need for conductive gel: in this case the electrodes are referred to as *dry*. Advances in material technology are producing surface electrodes that are more comfortable for consumer use, for example, electrodes embedded in flexible grids (Lapatki, van Dijk, Jonas, Zwarts, & Stegeman, 2004) or even embedded in fabrics (Paradiso, Loriga, & Taccini, 2004).

The typical biomedical analysis for diagnostic applications involves envelope detection, energy measurement (which relates the signal to physical force), and frequency characterization (DeLuca, 1997). Control applications generally involve signal acquisition from a number of deferential electrodes, feature extraction, and real-time pattern classification. The first examples of EMG-based real-time control systems were for prosthesis control and functional neuromuscular stimulation. Hefftner, Zucchini, and Jaros (1988), for example, report successful results from a system that can recognize two gestures generated from the shoulder and upper arm. The system must be specifically calibrated for each subject, and uses EMG signals from two channels.

Lukowicz, Hanser, Szubski, and Schobersberger (2006) presented a system based on wearable force-sensitive resistors to sense muscle activity. They showed a correlation between the mechanical deformation of the limb (measurable through force sensors placed on an elastic band adherent to the body) and muscle activity, especially fatigue. This approach allows the recording of activity that cannot be obtained through inertial sensors. Unfortunately, no sensing of pure isometric activity is reported. Strachan and Murray-Smith (2004) used accelerometers to measure muscle tremor as a form of isometric muscle activity. The system can detect the gestures of squeezing or holding a PDA in the user's hand, but requires individual calibration.

EMG for Human-Computer Interaction (HCI)

A number of studies have focused on EMG for users with physical disabilities (Coleman, 2001; Guerreiro & Jorge 2006). Putnam and Knapp (1993) developed a reconfigurable system to control generic graphical user interfaces. The system incorporates a continuous control mode where the contraction's amplitude is mapped to a parameter swing (sliders, scrollbars) and a gesture recognition mode that discriminates between two gestures and can be used for discrete selections. Gesture recognition is performed on a dedicated digital signal processing (DSP) board, is based on neural networks, and requires training for each user. Barreto, Scargle, and Adjouadi (1999) propose a system to control a mouse-like point–and-click interface using facial muscles. In addition to amplitude, the EMG signals' spectral features are analysed to increase performance. The system is not reported to require individual calibration for each user, and is implemented on a DSP board.

Other examples of EMG-based HCI include robotic control (Crawford, Miller, Shenoy, & Rao, 2005), unvoiced speech recognition (Manabe, Hiraiwa, & Sugimura, 2003), pointer control (Rosenberg, 1998), affective and emotional state recognition (Benedek & Hazlett, 2005; Healey & Picard, 1998), and a number of musical expression interfaces. For musical expression, the signal is used either in a continuous fashion, for example, with the amplitude being mapped to a variety of sound synthesis parameters, or through gesture recognition. The systems presented in this context are often wearable and allow movement of the performer on stage, yet they are not explicitly designed for the mobile everyday context. Knapp and Lusted (1990) present a generic battery-powered platform to control MIDI systems. Tanaka and Knapp (2002) complement EMG data with inertial sensor information, so that both isometric and isotonic activity can be monitored: muscle tension resulting in no motion and motion with constant muscle tension respectively. Dubost and Tanaka (2002) developed a wearable wireless musical controller supporting preprocessing of EMG signals and output interfacing with different standards (MIDI, RS232, and Ethernet), which requires calibration for every user.

Recent studies focus on the use of EMG for the recognition of an alphabet of discrete gestures. Fistre and Tanaka (2002) propose a system that can recognize six different hand gestures using two EMG channels on the forearm. The device is designed to control consumer electronics and is described as portable. Testing in a mobile context has not been reported. Wheeler and Jorgensen (2003) report the development and successful testing of a neuroelectric joystick and a neuroelectric keypad. Using EMG signals collected from four and eight channels on the forearm, they successfully recognize the movement corresponding to the use of a virtual joystick and virtual numeric keypad. Gestures mimicking the use of physical devices are successfully recognized using hidden Markov models. The system is proposed as an interface for mobile and wearable devices, but an embedded implementation is not reported, nor is testing in a mobile context.

SUBTLE GESTURES AND THE INTIMATE COMMUNICATION ARMBAND

The EMG signal's ability to detect isometric muscle activity (muscle activity not related to movement) allows the definition of a class of "subtle" or "motionless gestures." Motionless gestures are defined as specific, isolated, and volitional muscle contractions that result in little or no visible movement, and that are different from everyday muscle activation patterns. An example of a motionless gesture is a *brief contraction of the upper arm*, a gesture somewhat similar to a brief grasp of an object held in one's hand that can, however, be performed also with free hands. While it might be initially difficult for the reader to imagine such a contraction, minimal feedback about the gesture recognition makes it very easy to learn and to perform reliably, as demonstrated in a user study reported later in this chapter. In fact, the definition of this brief contraction was the result of a user-centred process also described later in the chapter.

Previous studies on the use of EMG for human computer interaction (mobile or not) do not explicitly consider subtlety, leading to a different approach. Tanaka and Knapp (2002) consider it a limitation that EMG cannot distinguish between muscle activity from movement and nonmovement. They remedy this by complementing EMG with inertial sensor (gyros) data in a multimodal fashion. Fistre and Tanaka (2002) and Wheeler and Jorgensen (2003) use EMG for hand-gesture recognition as an alternative to accelerometers or mechanical sensors for movement, but not for subtle gestures.

In addition to the emphasis on subtlety, the approach proposed here is different from other work on EMG (Fistre & Tanaka, 2002, Putnam & Knapp, 1993, Wheeler & Jorgensen, 2003) as it favours avoiding calibration or system training for each user, minimal computational complexity, and robustness against false positives in sacrifice of the variety of gestures recognized.

The Intimate Communication Armband was conceived as a generic input/output peripheral for mobile devices. It is worn on the upper arm invisibly under clothes (Figure 1), senses explicit subtle gestures, and provides localized tactile output. It connects wirelessly via Bluetooth to a phone or PDA, which can sit in the user's pocket or bag. Being a generic i/o device, it emits signals every time a gesture is recognized and accepts signals to activate the tactile display. In this way, complete freedom for a mapping strategy is left to the application designer.

The Intimate Communication Armband does not occupy the user's hands, and does not require hands to operate; hence, it is "hands free." On its own, it can be used for minimal communication and remote awareness: paired armbands can provide a very low bandwidth intimate and unobtrusive communication channel, if one vibrates

Figure 1. The Intimate Communication Armband can be made invisible by hiding it under clothing

every time the other one detects a motionless gesture. However, the device's greatest potential is realized when combined with a higher-resolution hands-free display to form a closed loop "hands free" system, such as with headphones, loudspeakers, or even high-resolution eyeglass displays. This can be highly advantageous in a number of everyday situations, for example, when the user is carrying objects, as well as for specific domains of applications, such as maintenance in which the users' hands are needed to perform a principal task, and the mobile computing system is used for remote guidance or accessing documentation, for example, an audio guide could be read through headphones and an armband could be used to advance, pause, or rewind the system. A tactile display can be used to give feedback about a subtle gesture being recognized, or it can deliver alerts and notifications.

Hardware

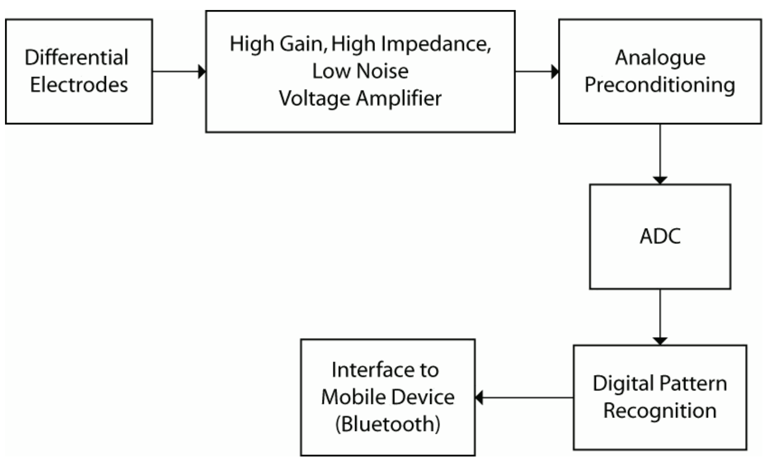
Custom hardware was developed to sense, amplify, and process EMG signals from the upper arm. Commercial EMG amplifiers are generally designed for biomedical applications, where high accuracy and reliability fully justify high prices. Moreover, such equipment is often used in controlled or semicontrolled hospital conditions, so devices are often worn on the belt or the patient's back, and connected through wires to electrodes on other body parts. While this setup allows more flexibility in electrode placement and multiple channel recording, it can be cumbersome to wear in everyday conditions. In contrast, cost for the intimate communication armband is below $100, the most expensive component being the Bluetooth module, which alone accounts for about half. Of course, the accuracy and reliability are not comparable with commercial biomedical devices, yet sufficient for the proposed application.

As detailed in Section 2, the EMG signal is a biopotential in the range of 100 μV to about 1 mV. The general system design for the subtle gesture sensor is illustrated in Figure 2, and it includes:

- Surface electrodes to pick up voltage signals on the body
- A signal amplifier and analog preconditioning stage
- An analog to digital converter
- A digital pattern recognition system
- An interface to applications on a mobile device or PC

The signal preconditioning was performed through analog rather than digital filters to keep the digital processing complexity low. This choice was made based on the detection algorithm's low computational cost, which can run on a low-power 8-bit RISC microcontroller.

Figure 2. Block diagram for the EMG subtle gesture recognition system



As compared to other sensing methods, EMG has had a number of practical difficulties, due to the need for contact electrodes and their placement (Rekimoto, 2001). However, EMG is worth studying, given its significant advantages in subtlety, and new developments in noncontact electrodes (Trejo, Wheeler, Jorgensen, Rosipal, Clanton, Matthews, et al. 2003) and smart materials.

The system includes two separate circuit boards to minimize interference: one for analog signal amplification and filtering, the other for digital processing and Bluetooth communication. The amplifier design is based on a portable electrocardiogram (ECG) sensor (Company-Bod & Hartmann, 2003). The system uses an integrated instrumentation amplifier in the first stage, with a right-leg driver-feedback stage to reduce noise. The right-leg driver feeds common mode signals back to the source, a design quite common for biosignal amplifiers (Webster, 1992). After the first stage, a first-order high-pass filter at 1.6 Hz is used to eliminate DC components, followed by a second-order Sallen-Key active low-pass Butterworth at 48 Hz with a gain factor of 10, for antialiasing and further noise reduction. A final stage with unity gain is used to offset the signal, centring it with respect to the analog to digital converter (ADC) range. An integrated voltage converter is used to provide +5 V and -5 V supply for the analog stage from a single cell 3.7 V, 130 mAH Li-Po battery. The circuit schematic is illustrated in Figure 3.

An Atmel 8-bit AVR microcontroller, the AT Mega168, is employed for analog to digital conversion, gesture recognition, and to drive the vibrating motor. The motor is driven through pulse width modulation (PWM) to allow fine tuning of the vibration intensity. The BlueGiga WT12 Bluetooth module is used for wireless communication, connected via a serial interface to the microcontroller. Another integrated voltage regulator is used to convert the battery voltage to 3.3 V, as required by the Bluetooth module. The board also includes a C-MOS driver and a protection diode for the vibrating motor, and two LEDs for displaying the microcontroller's status during debugging.

The two boards and the battery are housed in a box of about 3cm x 4cm x 2cm that is inserted into an elastic armband made for a commercial MP3 digital music player, as shown in Figure 4.

Figure 3. Circuit schematic for the EMG amplifier

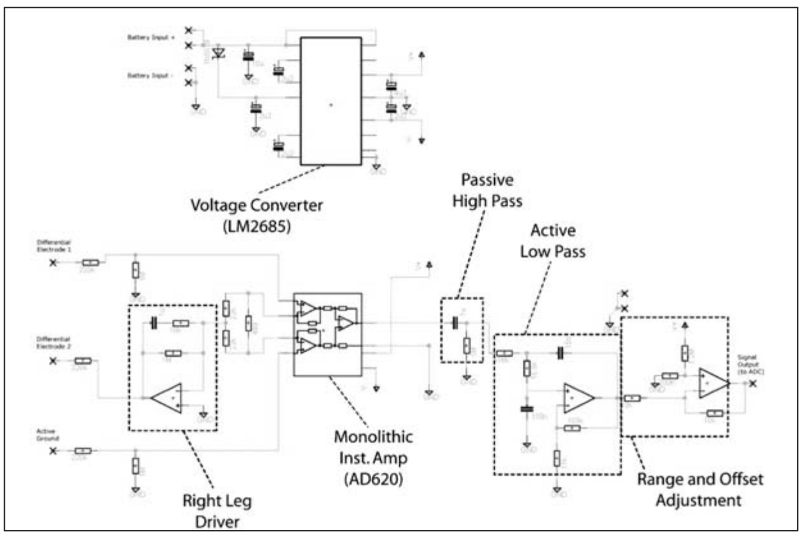
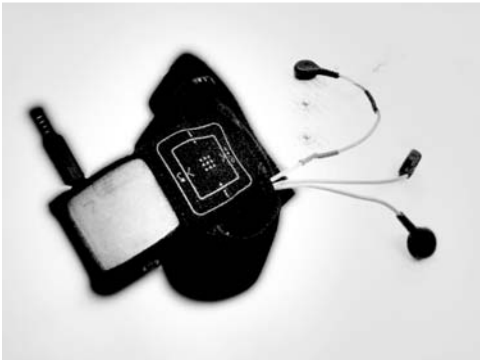


Figure 4. The second generation EMG sensor inside an armband holder for a commercial digital music player. The connector on the left of the photograph is used for recharging the battery and also as a power switch.



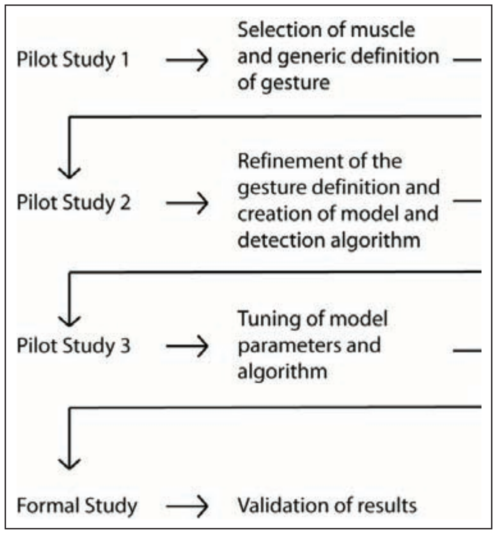
User-Centred Definition of Subtle Gestures

The design of the recognition algorithm and the definition of the gesture were done in parallel to satisfy two requirements: the gesture must be (1) natural for people to perform, and (2) different enough from "normal" muscle activity to avoid misclassification or "false positives." The gesture design was a user-centred iterative process. A number of exploratory, informal user studies were performed to ensure that the system would be natural and easy to use, as summarized in Figure 5.

The process started with a pilot study to select one muscle and subtle isometric contractions that fit the definition of motionless gestures. The test revealed the biceps as the best candidate because it lies superficially, making the signal fairly immune to activity generated by other muscles, and it is well defined, even in nonathletes. The gesture was defined as a brief contraction, such that it could be performed without being noticed, while the arm is unfolded, parallel to the body while the user is standing.

A second informal study was conducted to refine the definition of the subtle gesture and create a model and algorithm for its detection.

Figure 5. Outline of the design process for the subtle gesture recognition



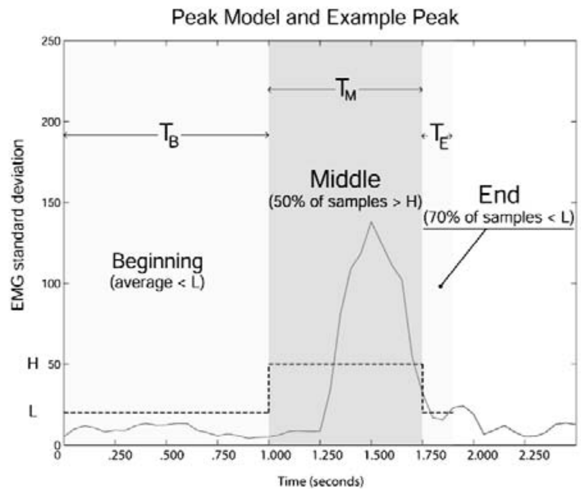
New subjects participated in the study and were chosen for a variety of muscle volumes. EMG signals were recorded from subjects performing the selected contraction, and compared with the signals generated by other types of muscle activity, such as moving in an indoor space, lifting objects of various weights, and gesticulating while talking.

The subjects were informed about the study's purpose, and the gesture was described to them in a not-detailed way (just as a "brief contraction of the biceps, i.e., the upper arm, that would not be very evident") so that they had some freedom in the way they performed it. This procedure aimed at exploring whether such a definition of "brief contraction" would be consistent across individuals, and to ensure that the gesture definition would be, to a certain extent, natural to perform, rather than defining a gesture a priori, and ask or force the users to learn it.

Subtle Gesture Model

The model resulting from the second study, depicted in Figure 6, is based on the standard deviation of the EMG signal, calculated with a sliding window of duration 0.2 s overlapping for

Figure 6. Model for the subtle gesture (dotted line) and an example gesture recording detected by the algorithm (solid line)



75% of its duration. A mathematical model and a recognition algorithm for the brief contraction were then created heuristically from observation of the data. A brief contraction was observed to correspond to a peak in the signal's standard deviation. Given the noise-like characteristics of the EMG signal (DeLuca, 1979), standard peak-detection techniques could not be employed. Rather, such peaks were modelled as follows: a "beginning" interval of duration T_B of low activity ("silence"), followed by a "middle" interval of high activity of duration T_M and then again, low activity for an "end" interval of duration T_E. High activity and low activity were defined respectively as the signal's standard deviation being above a threshold H and below a threshold L. To allow some tolerance in the model, the condition on the history is imposed on an average of its values; the condition on the middle needs to be satisfied by 50% of the samples, and the condition on the end by 70% of the samples. To increase the resilience to false positives caused by motion artefacts, a zero-crossing counter is included in the detection algorithm to reject low-frequency components.

The model definition is stricter on the contraction's duration than it is on the gesture's intensity. This is because the preliminary study showed that the duration was more consistent than the intensity across users, despite the fact that no specific indication to users was given about either. One disadvantage of this model is it requires a complete gesture before the recognition can take place. The recognition could be made faster by removing the "end condition" for the gesture's closure; however, this would cause an increase in false positives.

The tuning of the model's five parameters required a third informal study. New and returning users were informally asked to test the system. The testing was conducted to stress the system to produce false positives and false negatives. The iterations continued until the number of false positives approached zero and the system recognized contractions performed by any user.

Two Gestures: Long and Short

Once the recognition worked robustly on one gesture, a two-gesture alphabet was explored. The gestures were defined as two short, subtle contractions of different durations. This corresponded to varying the middle interval T_M's duration together with its tolerance. The results obtained at this point were then validated with the first formal user study, described next below.

EVALUATION

Three user studies were performed to validate the design of the EMG-based interaction technique (Costanza, Inverso, & Allen, 2005; Costanza, Inverso, Allen, & Maes, 2007). The novelty of the approach required the evaluation of several different aspects of the interaction, including ease of learning, correct gesture recognition rate, amount of information that can be expressed through subtle gestures, usability of the gestures within a realistic multimodal interface, and noticeability of the interaction to others. Because of the unnoticeable nature of the interaction, it was impossible to define Wizard–of-Oz type studies; therefore, all experiments were conducted using working prototypes of the gesture sensor.

First Study: Learning and Recognition Rate

The first study had three main objectives: (1) to assess whether subjects could learn how to perform the gestures without training, simply by trial and error, receiving minimal feedback; (2) to measure the recognition rate of subtle gestures through the algorithm described; (3) to test whether multiple gestures could be defined on a single muscle.

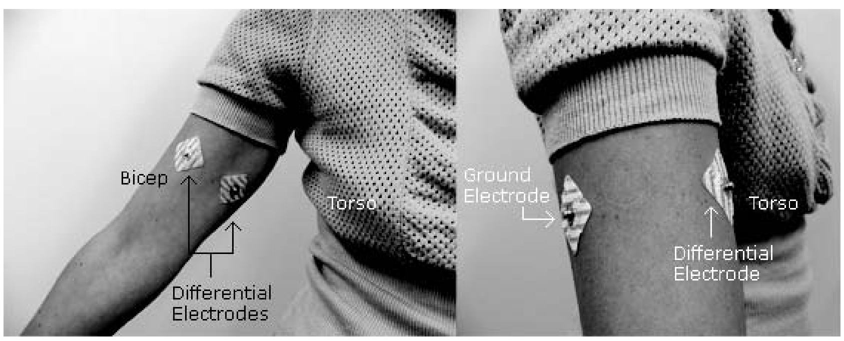
The experiment was carried out in a simulated mobile scenario: subjects were asked to perform experimental tasks with the device while walking around obstacles in a trafficked walkway in the Media Lab Europe. The setup was similar to one reported by Pirhonen, Brewster, and Holguin (2003), who noted that this mobile context allows us to "take measurements of the usage of the device whilst the users were mobile but was not as formally controlled as a laboratory study, which would lack realism and ecological validity." Subjects were asked to wear the EMG-sensor armband and a pair of wireless headphones so that they could receive auditory cues and feedback while being free to move around. An experimenter applied disposable, solid-gel, self-adhering, Ag/AgCl 9-mm disc surface electromyogram electrodes in three positions around the upper arm of each subject's dominant hand, as illustrated in Figure 7. To ensure signal quality the participant's skin was prepared with an abrasive gel before the electrodes' application.

Participants were 10 adults, 5 women and 5 men, ages 23 to 34, all colleagues from Media Lab Europe, who volunteered to take part in the study. All were naive in that they had not used an EMG-based interface before, with the exception of subject 8, who had taken part in a pilot study.

At the beginning of the experiment, subjects were given written instructions informing them that the study was assessing EMG as a subtle interface for mobile devices, and that the system would recognize brief contractions of the upper arm. The instructions specified that the contraction recognized has a minimum and maximum duration and a minimum strength requirement.

To test how easy the gestures are to learn, subjects were invited to familiarize themselves with

Figure 7. Electrode placement used for the first user study



the system until they could comfortably control it. In this phase, participants stood and only heard auditory feedback when the system recognized a contraction. No coaching or further feedback as to the contraction's amplitude or duration was given to the participants; so they were unaware of why the algorithm was or was not recognizing the contraction. They were only aware if the contraction was recognized. If participants did not comfortably control the device within 15 minutes, they were given further feedback by an experimenter who could observe the recorded EMG signals.

After the initial familiarization, participants were asked to engage in the first experimental task: perform a gesture every time they heard an audio cue through the headphones, while walking around the obstacles. The same auditory feedback, as confirmed when a contraction was recognized. Subsequently, subjects were asked to repeat the task three more times, with variations on the gesture duration: in one case, they were asked to always try and perform gestures that were as short (in time) as possible while still being recognized by the system; in another case, to always perform gestures as long as possible (but still recognized by the system); and finally, to perform a mix of "short" and "long" gestures in response to different auditory stimuli. The task with only "long" gestures and the task with only "short" gestures were run in counterbalanced order, so that half of the subjects performed "long" gestures first and the other half "short" gestures first. Each task was preceded by a brief familiarization on "short" or "long" gestures in which again, participants stood and only heard an auditory feedback when the system recognized a contraction. In all cases, the same real-time detection algorithm was used across participants without calibration or modification, and it recognized contractions of duration 0.3 to 0.8 seconds. Therefore, the exact definition of "short" and "long" gestures was left to the individual.

Results and Discussion

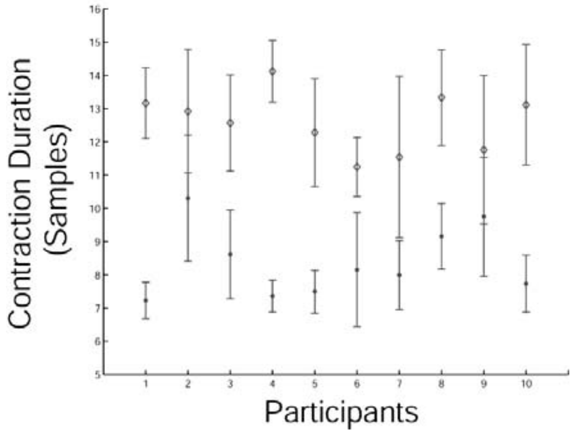
The online recognition rates for the four contraction walking tasks were generic 96%, short 97%, long 94%, and mixed 87%. No false positives were detected while online during the first walking task. This accuracy level indicates that EMG-based motionless gesture recognition can be used successfully to control a mobile interface.

In the first familiarization task, participants were able to control the system in an average of 3.75 minutes (SD=2.17), excluding the three participants who reached the 15 minute time limit and required additional feedback. The participants who received feedback (2, 9, and 10), all had the same difficulty that their contractions were too long. They were told, once, to make their contractions shorter, and then they were able to control the system in 11.75, 1.78, and 5.48 minutes, respectively.

Off-line analysis was performed on the data from the short- and long-contraction walking tasks to determine if short and long contractions are separable into two gestures for control. Figure 8 shows the mean and standard deviations for the short and long contraction durations. From the data, a duration boundary of 0.5 seconds was used to create a new recognition algorithm that recognized long and short contractions separately. As with the original algorithm, only the first recognition was counted; any additional recognitions were ignored until the next stimuli. Applying this new short-long detection algorithm to the mixed contraction data resulted in an overall accuracy of 51%, with 55% shorts recognized and 47% longs recognized. The misclassification rate for shorts as longs was 33%, and the misclassification rate for longs as shorts was 11%.

The off-line recognition of short and long contractions using the mixed data set was fairly low. This may have occurred because the online algorithm recognized a small range of contraction durations; therefore, the longs may not have been sufficiently different from the shorts for the participants to accurately produce them. The contraction duration's range was set from pilot studies, which indicated that very long muscle contractions cause most false positives; therefore, a trade-off between reproducibility of long and short contractions and increased false positives can occur if the range is widened.

Figure 8. Mean and standard deviation error bars for long and short contraction durations; closed circles indicate means for short and open circles indicate means for long.



It is important to note the durations of the short and long contractions are subjective because the participants were not given feedback about their actual durations. Therefore, the participants trained themselves on what they considered were long and short contractions. If the participants were given feedback on their contraction durations, they may learn to consistently make different long and short contractions.

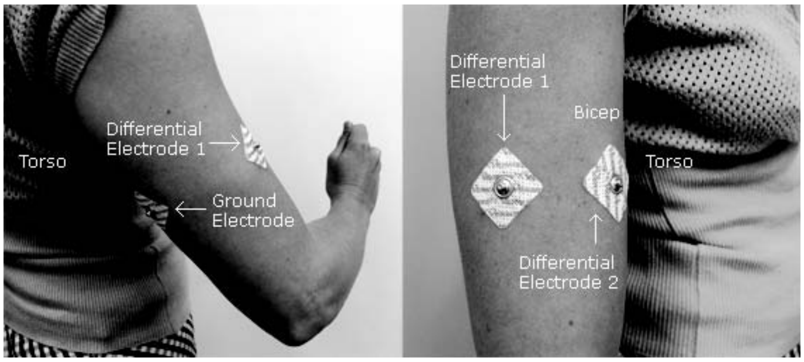
After the experiments, some participants stated that they felt longs were more difficult than shorts. In addition, it was noticed that the three participants that required feedback in the first familiarization task became frustrated when they could not make the system recognize their contractions; however, by the end of the experiment they were comfortable using the system.

Second Study: Multimodal Realistic Interaction

Once the first study confirmed the basic functionality of the system, a second study was performed to explore usability in more realistic conditions: subtle gestures were used to select one of four items of an audio-menu. The task required expression of multiple bits of information through subtle gestures, two strategies to achieve this were compared, either by using a time-based interaction or by using two armband devices at the same time. In one condition, defined as *one-arm*, the menu items were iteratively scanned, so that the contraction of just one arm could be used to select the current item. In the other, *two-arm*, condition gestures from one arm were interpreted as “next’”and used to advance through the menu items, while gestures from the other arm were used to select the current item.

Two tasks were defined; in each task subjects had to perform a selection in response to a number of cues using one of the conditions described. Wireless headphones were used to display the audio menu, read by a synthetic voice (AT&T, 2004), and deliver audio cues. The cues mimicked incoming phone calls from four callers: each cue consisted of a synthetic voice (the same one used for the menu) announcing “Incoming call from...” followed by the caller’s number or name. After each announcement, subjects could access the audio menu and select one item. Subjects were instructed to select a specific menu item in response to each of the four callers. Similar to the first study, participants performed each task while navigating eight-meter laps around obstacles in a regularly trafficked walkway in the MIT Media Lab. Each of the two tasks was preceded by a short familiarization session. All subjects participated in both tasks: within-subjects design and the tasks were performed in fully counterbalanced order.

Figure 9. Electrode placement used in the second user study



The subject's walking speed during each task was measured as an index for the interface's effectiveness. Petrie, Furner, and Strothotte (1998) pointed out that if a mobile interface has a negative effect on users, it will be reflected in them slowing down while walking. The same measure was later used in other mobile HCI studies (Lumsden & Brewster, 2003; Pirhonen et al. 2003). The subject's preferred walking speed (PWS), that is, the speed at which they walk while not using any mobile device, was measured at the experiment's beginning as a comparison.

Participants were 12 adults; 8 women and 4 men, all volunteers recruited through posters on the MIT campus and university mailing lists. All expressed interest to participate in the study via e-mail, demonstrating a minimum familiarity with computer systems, and they were compensated $10 per hour. All subjects were naive in that they had not used an EMG-based interface before.

Subjects were asked to wear one or two armband devices, depending on the task. Similar to the first experiment, electrodes were placed around each of the subjects' upper arms; however, in this experiment, the participant's skin was not abraded, as an improved amplifier eliminated signal artefacts due to skin creams or lotions. A new electrodes position, illustrated in figure 9, was chosen to avoid the artefacts caused by electrodes pressing against the torso, noticed for some participants in the first study.

Results and Discussion

Overall, subjects performed correct selections of items from the audio menu for 226 of the 235 stimuli presented, corresponding to 96.2% correct selections. Incorrect selections were performed in six cases (2.5%); in all except one of these, an item adjacent to the correct one was selected. In three cases (1.3%) no selection was made. In the two-arms condition, subjects performed correct selections for 120 of the 123 stimuli presented, 97.6% correct; in the same condition, two erroneous selections (1.6%) and only one missed selection (0.8%) occurred. In the one-arm condition, subjects performed correctly for 106 of the 112 stimuli: 94.6%. The number of errors in this condition was four (3.6%) and two misses (1.8%) occurred. Out of the 12 subjects, 7 performed perfectly in both conditions (100% correct selections), while 2 subjects achieved a perfect score on at least one condition. Only five false positives were detected during the entire experiment, but these did not affect the task performance as they happened after a selection was made and before the subsequent stimulus. Additionally, two times subjects reported that an incorrect selection (included in those reported above) was from a false positive.

The high overall accuracy indicates that EMG can, in general, be used successfully in complex and multimodal interfaces. The performance was high in both conditions, demonstrating that the interface bandwidth can be improved either by using controllers on multiple muscles or by using time-based selection strategies. The higher percentage of correct selections in the *two-arm* condition, and the preference expressed by the subjects, suggest that this interaction modality is more efficient than the other one, of course with the extra expense of an additional controller.

A one-way ANOVA showed no significant differences in the subjects' walking speed corresponding to different tasks. Most of the subjects walked slower when operating the interface, however, four subjects walked faster in the *two-arms* conditions than when they were walking without interacting with the device, and three subjects walked faster in the *one-arm* condition than when not operating the interface. These results suggest that controlling an EMG-based interface, with one or two arms, does not involve high workload, nor does it require a high amount of attention (Lumsden & Brewster, 2003; Petrie et al. 1998, Pirhonen et al. 2003). However, further research is required for more conclusive findings.

Eight of the 12 subjects learned to control the device very quickly, and 4 naturally performed the gesture without much arm movement. When asked at the end of the experiment, 7 of 10 subjects expressed a preference for the *two-arms* condition, generally because this provided more control and faster operation; only 2 of 10 subjects preferred the *one-arm* condition, 1 did not express a preference. Most of the subjects spontaneously reported that they enjoyed taking part in the experiment and experienced a novel and unusual way to control a computer interface.

Third Study: Assessing Noticeability

One of the strongest motivations for the use of EMG in the context of mobile HCI is the ability to sense isometric muscular activity, which enables the creation of input interfaces that are subtle, unobtrusive, and unnoticeable by those around the users. An experiment was performed to formally assess how noticeable these gestures are.

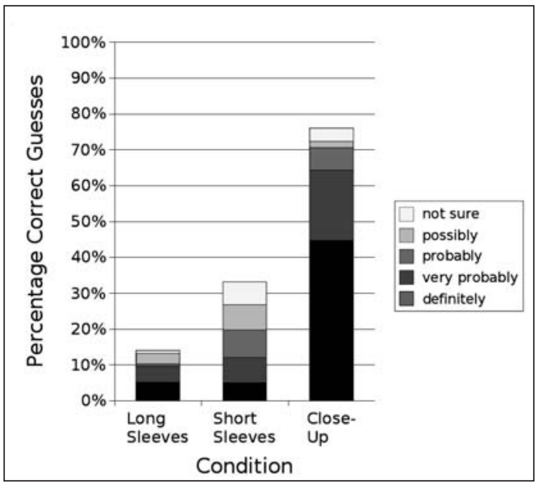
The same subjects who took part in the second study were asked to watch a video recording of a trained user activating the interface, and to try and guess when a gesture was performed. The experiment was performed immediately after the completion of the previous one, so all subjects were familiar with the EMG-based interface. The video showed an actor performing subtle gestures with his right upper arm while talking with someone off screen. The recording had no audio and it was divided into three scenes: a medium shot of 135 seconds with the actor wearing long sleeves; a shot of 144 seconds with the same framing and

Figure 10. GUI used to rate the EMG video

the actor wearing short sleeves; and a close up of the arm with the electrodes and the armband with duration of 41 seconds. The actor was really activating an armband device, whose recognition output was used as ground truth.

The video was shown on a standard 17’’ LCD computer display in a video window that measured approximately 8’’ by 6’’ at the centre of the screen. Under the video window the sentence “The person in the video is activating the EMG interface” and five buttons labelled “definitively,” “very probably,” “probably,” “possibly,” “not sure” were displayed, as illustrated in Figure 10. Subjects were informed about the purpose of the study and instructed to click on one of the five GUI buttons every time they believed the person in the video was performing a gesture recognizable by the EMG-based system.

Figure 11. Results of video rating in the second user study



Results and Discussion

Guesses were considered correct if they were within 1.5 seconds of ground truth. For the long sleeves video section, subjects correctly guessed when a contraction was performed for only 13.9% of the attempts (19 correct guesses over 137 attempts). For the short sleeves section, 33.1% of the attempts were correct (47 correct guesses over 142 attempts). Finally, in the close-up case, 75.9% of the attempts were correct (85 over 112 attempts). The subjects’ confidences on correct guesses are reported in Table 1. The same data is visualized in Figure 11.

The results show that subjects cannot easily guess when the interface is activated, confirming that the interface is subtle. In the experiment, subjects were told that the person in the video would, at some point, activate the interface; in reality this information would not be available, making the chance of noticing the interface even smaller. Most of the subjects informally reported that they found it difficult to tell when a contraction was performed.

The results can be compared to the probability of a correct uninformed guess, that is, the probability of guessing correctly, assuming that subjects did not look at the video and guessed randomly. This situation can be modelled with the attempts having a uniform random distribution. Considering each “long sleeves” and “short sleeves” sequence separately, and remembering that an attempt is considered correct if it is within 3 seconds of a contraction, a high enough number of attempts evenly spaced in time would give a 100% chance of guessing correctly. The minimum number of attempts for a 100% chance of guessing

Table 1. Video rating results in the third user study

	% overall correct	definitely	very probably	probably	possibly	not sure
Long Sleeves	13.9%	36.84%	31.58%	5.26%	21.05%	5.26%
Short Sleeves	33.1%	14.89%	21.28%	23.40%	21.28%	19.15%
Close-Up	75.9%	58.82%	25.88%	8.24%	2.35%	4.71%

is $N_{100\%} = D_s / D_a$, where D_s is the duration of the sequence and D_a is the uncertainty interval, in this case, 3 seconds. In the "long sleeves" condition, D_s is 135 seconds, so $N_{100\%} = 45$ attempts would give a 100% chance of guessing correctly.

During the experiment, subjects cumulatively attempted to guess 137 times, corresponding to an average of 11.4 attempts per subject, and to an 11.4 / 45 = 25.3% chance of correctly guessing. In the "short sleeves" condition, 142 attempts were made corresponding to an average of 11.8 attempts per subject, over 144 seconds, so $N_{100\%} = 144 / 3 = 48$ and the uninformed guess chance is 11.8 / 48 = 24.6%.

Therefore, in the long sleeves condition, the subjects guess performance, 13.9%, was much worse than completely random, 25.3%, implying that watching the video did not help guessing, confirming that the contractions are unnoticeable. In the short sleeves case, subjects guessed 8.5 percentage points better than chance; however, overall fairly low. In the close-up condition, subjects guessed correctly most of the time.

General Discussion

The results from the two studies demonstrate that the Intimate Communication Armband can be reliably used in a mobile context. In both the initial and the audio menu experiments, novice subjects learned to use the system very quickly, with little feedback about their performance. Subtle gestures proved to be effective in controlling a multimodal interface even when mobile. Although expressing different subtle gestures with a single arm seems not to be very reliable (at least with the current detection algorithm), subjects did not have problems in using multiple muscles at the same time, nor to use a single muscle to select one of many options presented over time.

The gestures recognized by the armband device are indeed subtle; the last experiment's results showed that it is hard for observers to guess when someone is performing a gesture.

FUTURE TRENDS

Further investigation should explore the use of more advanced analysis techniques for the detection of subtle gestures, such as autoregressive modelling, which has been reported to be successful in some EMG literature (Hefftner et al, 1988). The performance of subtle gestures while users are engaged in tasks that occupy their hands or involve specific movements of their arms should be formally investigated, focussing not only on the physical challenges, but also on the cognitive demands related to performing different motor tasks. To improve the device's comfort, dry electrodes or electrodes embedded in fabric (Paradiso et al., 2004) should be included in the armband design. The potential of localized tactile cues should be explored: tactile stimuli on different parts of the body can convey a large amount of information. Armbands should be paired so that one vibrates when the other recognizes a subtle gesture to form a simple intimate communication system for remote awareness. The device should also be integrated within specific mobile applications, such as browsing audio documentation or navigation guidance. Higher-level evaluation of these applications should analyse how users adopt it for day to day use.

Generally, there are many opportunities to develop biosignal processing techniques that run on embedded devices and the exploration of new application domains, including, for example, games.

CONCLUSION

This chapter has shown that an EMG-based wearable input device, the Intimate Communication Armband, can be effectively employed in a mobile context for subtle and intimate interaction. The device detects *subtle motionless gestures*: explicit muscle contractions resulting in little or no movement. Experimental results show that the gestures are reliably recognized without training, neither for the recognition algorithm nor for users. Subjects were able to reliably control an audio interface using one or two arms while engaged in a walking task without problems. An experiment designed to evaluate the subtlety of the interface revealed that it is very difficult for observers to guess when a trained user is performing subtle gestures. The armband device also includes a tactile display, based on a vibrating motor, and can be made invisible by being worn under clothes.

The design of interfaces and interaction techniques for mobile devices should take into account social acceptance and allow devices to be active but not disruptive. The construction and evaluation of the prototype proposed in this chapter demonstrates that it is possible to realize usable mobile interfaces that are intimate and subtle and therefore, socially acceptable.

ACKNOWLEDGMENT

This research was initiated at Media Lab Europe and continued at MIT Media Lab through the support of the Things That Think (TTT) and the Digital Life (DL) consortia. The authors would like to acknowledge Alberto Perdomo and Juanjo Andres Prado for hardware design and support. Our gratitude goes to Ian Oakley for invaluable suggestions on the user studies design. We are also thankful to Joe Paradiso, Jim Barabas, David Bouchard, Mark Feldmeier, David Merrill and Sajid Sadi at the Media Lab for their insights.

REFERENCES

AT&T. (2004). AT&T natural voices text-to-speech engine. Retrieved from http://www.naturalvoices.att.com/products/tts_data.html

Barreto, A., Scargle, S., & Adjouadi, M. (1999). Hands-off human-computer interfaces for individuals with severe motor disabilities. *HCI (2)* (pp. 970-974).

Benedek, J., & Hazlett, R. (2005). *Incorporating facial emg emotion measures as feedback in the software design process.* Technical report, Microsoft Corporation. Johns Hopkins University School of Medicine.

Coleman, K. (2001). Electromyography based human computer-interface to induce movement in elderly persons with movement impairments. In *WUAUC'01: Proceedings of the 2001 EC/NSF workshop on Universal accessibility of ubiquitous computing* (pp. 75–79). New York, NY: ACM Press.

Company-Bod, E., & Hartmann, E. (2003). Ecg front-end design is simplified with microconverter. *Analog Dialogue*, *37*(11).

Costanza, E., Inverso, S. A., & Allen, R. (2005). Toward subtle intimate interfaces for mobile devices using an EMG controller. In *CHI '05: Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 481-489). New York, NY: ACM Press.

Costanza, E., Inverso, S. A., Allen, R., & Maes, P. (2007). Intimate interfaces in action: Assessing the usability and subtlelty of EMG-based motionless gestures. In *CHI '07: Proceedings of the SIGCHI conference on Human factors in computing systems*. New York, NY: ACM Press.

Crawford, B., Miller, K., Shenoy, P., & Rao, R. (2005). Real-time classification of electromyographic signals for robotic control. In *AAAI* (pp. 523–528).

DeLuca, C. J. (1979). Physiology and mathematics of myoelectric signals. *EEE Transactions on Biomedical Engineering*, *26*, 313-325.

DeLuca, C. J. (1997). The use of surface electromyography in biomechanics. *Journal of Applied Biomechanics*, *13*(2), 135-163.

Dubost, G., & Tanaka, A. (2002). A wireless, network-based biosensor interface for music. In *Proceedings of International Computer Music Conference (ICMC).*

Fistre, J., & Tanaka, A. (2002). *Real time emg gesture recognition for consumer electronics device control*.. Retrieved from http://www.csl.sony.fr/ atau/gesture/

Fortunati, L. (2002). The mobile phone: Towards new categories and social relations. *Information, Communication & Society*, *5*(4), 513-528.

Goffman, E. (1963). *Behavior in public places*. The Free Press.

Guerreiro, T. J. V., & Jorge, J. A. P. (2006). EMG as a daily wearable interface. In *Proceedings of the First International Conference on Computer Graphics Theory and Applications (GRAPP 2006)*, Setúbal, Portugal.

Hansson, R., Ljungstrand, P., & Redström, J. (2001). Subtle and public notification cues for mobile devices. In *Proceedings of UbiComp 2001*, Atlanta, Georgia, USA.

Healey, J., & Picard, R. W. (1998). Startlecam: A cybernetic wearable camera. *ISWC* (pp. 42-49).

Hefftner, G., Zucchini, W., & Jaros, G. (1988). The electromyogram (emg) as a control signal for functional neuromuscular stimulation-Part I: Autoregressive modeling as a means of emg signature discrimination. *IEEE Transactions on Biomedical Engineering*, *35*(4), 230-237.

Knapp, B., & Lusted, H. (1990). A bioelectric controller for computer music applications. *Computer Music Journal*, *14*(1).

Lapatki, B. G., van Dijk, J. P., Jonas, I. E., Zwarts, M. J., & Stegeman, D. F. (2004). A thin, flexible multielectrode grid for high-density surface emg. *J Appl Physiol*, *96*(1), 327-336.

Lukowicz, P., Hanser, F., Szubski, C., & Schobersberger, W. (2006). Detecting and interpreting muscle activity with wearable force sensors. In *Proceedings Pervasive 2006*.

Lumsden, J., & Brewster, S. (2003). A paradigm shift: Alternative interaction techniques for use with mobile and wearable devices. In *CASCON '03: Proceedings of the 2003 conference of the Centre for Advanced Studies on Collaborative research* (pp. 197-210). IBM Press.

Manabe, H., Hiraiwa, A., & Sugimura, T. (2003). Unvoiced speech recognition using emg - Mime speech recognition. In *CHI '03 extended abstracts on Human factors in computing systems* (pp. 794–795). New York, NY: ACM Press.

Marti, S., & Schmandt, C. (2005). Giving the caller the finger: Collaborative responsibility for cellphone interruptions. In *CHI Extended Abstracts* (pp. 1633-1636).

Okabe, D., & Ito, M. (2005). Keitai and public transportation. In M. Ito, D. Okabe, & M. Matsuda (Eds.), Personal, portable, pedestrian: Mobile phones in Japanese life. Cambridge: MIT.

Oulasvirta, A., Tamminen, S., Virpi, R., & Kuorelahti J. (2005). Interaction in 4-second bursts: The fragmented nature of attentional resources in mobile HCI. In *CHI '05: Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 919-928). New York, NY: ACM Press.

Paradiso, G., Loriga, G., & Taccini, N. (2004, August). Wearable health care system for vital signs monitoring - Medicon 2004 conference. In *Proceedings of MEDICON 2004*.

Petrie, H., Furner, S., & Strothotte, T. (1998). Design lifecycles and wearable computers for users with disabilities. In *Proc. First workshop on HCI mobile devices*. Glasgow, UK.

Pirhonen, A., Brewster, S., & Holguin, C. (2003). Gestural and audio metaphors as a means of control for mobile devices. In *CHI '02: Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 291-298). New York, NY: ACM Press.

Putnam, W., & Knapp, B. (1993, June). The use of the electromyogram in a man-machine interface. In *Proceedings of the Virtual Reality and Persons With Disabilities Conference*.

Rekimoto, J. (2001). Gesturewrist and gesturepad: Unobtrusive wearable interaction devices. In *ISWC '01: Proceedings of the 5th IEEE International Symposium on Wearable Computers* (p. 21). Washington, DC: IEEE Computer Society.

Rosenberg, R. (1998). The biofeedback pointer: Emg control of a two dimensional pointer. In *ISWC* (pp. 162–163).

Strachan, S., & Murray-Smith, R. (2004). Muscle tremor as an input mechanism. In *Proceedings of UIST 2004*. New York, NY: ACM Press.

Tanaka, A., & Knapp, R. B. (2002). Multimodal interaction in music using the electromyogram and relative position sensing. In *Proceedings of New Interfaces for Musical Interaction (NIME)*. Medialab Europe.

Trejo, L., Wheeler, K., Jorgensen, C., Rosipal, R., Clanton, S., Matthews, B., Hibbs, A., Matthews, R., & Krupka, M. (2003). Multimodal neuroelectric interface development. *Neural Systems and Rehabilitation Engineering, IEEE Transactions on*, *11*(2), 199-203.

Webster, J. G. (1992). *Medical instrumentation: Application and design (*2nd ed.). Houghton Mifflin.

Wheeler, K. R., & Jorgensen, C. C. (2003). Gestures as input: Neuroelectric joysticks and keyboards. *IEEE Pervasive Computing*, *2*(2), 56-61.

KEY TERMS

Electrode: Electrically conductive element placed as close as possible to a signal source (the muscle for EMG signals) acting as the interface between the body and the signal recording apparatus

Electromyographic signal or EMG Signal: An electrical voltage signal generated by muscle activity

Intimate Communication Armband: Wearable input/output device capable of detecting subtle gestures from the upper arm and of delivering tactile cues

Intimate Interfaces: Discrete interfaces that make interaction with mobile devices private and concealed as much as possible, in order to minimize the disruption of colocated individuals

Isometric Muscle Activity: Muscular contraction that does not produce movement; that is, the muscle length is constant; for example, pushing against a wall.

Subtle Gesture or Motionless Gesture: Voluntary muscle contractions that result in little or no visible movement—based on isometric muscle activity

Surface Electrode: Conductive metal plates, typically composed of Ag/AgCl metal, placed on the skin surface and kept in place through adhesive or elastic bands. Sometimes conductive gel is used to improve the electrical conduction between the skin and the electrode.

ENDNOTE

1 Impedance is the resistance to current flow. If the impedance between the electrode and skin is high, the muscle's electrical activity will not be conducted through the electrodes properly.

Chapter XXXII
Mobile Camera-Based User Interaction

Tolga Capin
Bilkent University, Turkey

Antonio Haro
D4D Technologies, USA

ABSTRACT

This chapter introduces an approach for user interaction on mobile devices, focusing on camera-enabled mobile phones. A user interacts with an application by moving their device, and the captured camera video is used to estimate phone motion or interact with the real world. We first survey technical issues, recent research results, and then present a prototype implementation and discuss various ways how phone motion can be used for different tasks, such as navigating through large number of media files, and phone motion and shake detection for gaming. The results and discussion may guide interface designers when targeting camera-based user interfaces.

INTRODUCTION

Mobile devices currently support key-modal interfaces through joypad/direction keys and numerical keyboard. On devices with larger form-factors, additional keys provide a better user experience for complex tasks such as navigating through large amounts of content, since keys can be dedicated to specific tasks such as page-up/down and choosing zoom level. Smart phones cannot easily make use of such keys due to limited physical space. Stylus-based interaction with touch-sensitive screens has emerged as an alternative, but it requires two-handed interaction, and has been shown to cause additional attentional overhead in users.

Consequently, alternative interaction techniques are desired. Physical sensors have been added to mobile devices for user interaction, such as accelerometers (Hinckley, Pierce, Sinclair, & Horvitz, 2000), but these can be difficult to integrate into existing consumer-level devices at both the software and hardware level at a low cost. In addition, such sensors are known to have error buildup over time, since some infinitesimal acceleration is always measured.

As another alternative, over the recent years, a number of solutions have been proposed for using the camera as the input device, where incoming video is used to estimate phone motion and to interact with the user's physical environment. These approaches provide a more direct user interaction maximizing the use of the display, minimizing attentional overhead to the user, and permitting one-handed interaction. With camera-based interaction, the user points directly on objects or changes their view by moving the phone. The user is provided with a means of navigating and manipulating individual objects, each of which has a direct display representation. The user applies actions directly to their view or to the objects by selecting them.

There are many application scenarios that could take advantage of camera-based user interaction. For general interaction with the device, the user can be provided a number of camera-based interaction primitives, such as gestures for scrolling and selecting. In games, users can control their viewpoint in the 3-D environment by physically moving their phone around in the real world to look up/down/left/right in the game world. In physical user interfaces, tags can be placed in real world, which can be scanned by the mobile device.

In this chapter, we survey recent research results, survey how camera based UIs can be used for different tasks and applications, and present a prototype implementation.

UNDERSTANDING COMPUTER VISION TECHNOLOGIES

Towards the goal of building applications that support camera-based interaction, a computer vision framework is needed on handheld devices. Computer vision is a large part of camera-based user interaction, and its limitations should be understood for designing interfaces; therefore, we discuss its main issues as the first step in this direction.

The mobile computer vision features required are a subset of the functionality on desktops. There are significant limitations on mobile devices, however:

- Mobile device CPUs have been limited in computing power. Only recently, high-end phones have started to support floating-point units. Hence, the use of floating point computations has to be minimized.
- The optics and image sensor chips of integrated cameras in mobile devices are targeted for imaging and video capture applications and as a result, provide limited quality for image processing tasks. For example, in smart phones, camera calibration is significantly more difficult due to the large amounts of lens distortion present. Such distortion makes the calibration algorithms commonly used in PCs infeasible for smart phones, as the recovered parameters will not be accurate enough for detecting 3-D position.
- Mobile devices do not support modifying camera focus and fixed focus cameras can only be effective in a certain depth range.
- In mobile devices, signal noise is more prevalent. Noise is caused by many factors, for example, bad camera lenses used, electronic noise caused by CCD camera, and "algorithmic" noise introduced by the imaging chain (e.g., white balance correction, exposure, gamma correction, color, shading, geometrical, noise-reduction). Images also have low contrast, varying brightness, and blurred edges.
- Battery power is a major consideration, limiting the type of applications and context that can be supported (e.g., it is not possible to have an always-on interaction scenario).

Correctly interpreting the observed motion of the objects or the global motion of the camera from video requires accurate tracking. To determine the motion direction, various tracking algorithms have been proposed. The solutions proposed in the literature can be divided into two main categories: markerless and marker-based techniques. Markerless tracking-based solutions analyze the video and detect important features, such as edges, corners, or corner-like features; or use motion-flow techniques for the global motion of the camera. Marker-based solutions use a visual tagging system that is based on printed 2-D markers that are placed in the environment and identified by mobile cameras. We describe the two tracking approaches in the next section.

Markerless Tracking

Markerless tracking systems do not assume any presence of a known object or a structure in the capture video. Thus, they provide a more general solution than marker-based systems, described later. Two different approaches are possible to achieve markerless tracking in mobile devices: template matching and optical flow.

- **Template-matching-**based solutions use an image region to track. These regions can be a rectangular block in the video (Figure 1), or arbitrary shapes. Although these systems have reliability issues, speed is their major advantage. The systems can use larger windows to capture more motion, but more processing is needed. Template matching is simple to implement, but requires good features to track.
- **Optical-flow-**based solutions are based on calculating the direction and motion speed of the features in the image, using the velocity field of pixels between two frames. The entire image can be used for tracking, increasing the correctness of the solution. However, the disadvantages of these methods are that the vector field may not be smooth (due to pixel disagreements) and the assumption of constant brightness is not always correct.

Mobile-camera-based markerless tracking has been researched by several groups:

- Rohs (2004) perform tracking based on dividing incoming camera frames into blocks, and then determining how the blocks move given a set of discrete possible translations and rotations.
- Haro et al. (Haro, Mori, Capin, & Wilkinson, 2005) propose a solution based on tracking individual corner-like features observed in the entire incoming camera frames. This allows the tracker to recognize sudden camera movements of arbitrary size, as long as at least some of the features from the previous frame are still visible, at the trade-off of not detecting rotations.
- Hannuksela et al. (Hannuksela, Sangi, & Heikkila, 2005) propose a region-based matching approach, where a sparse set of features are used for motion analysis, to-

Figure 1. Motion simulation. A base image with sliding window locations on the left (a) and a frame with ground truth motion (b) on the right.

(a) *(b)*

gether with a Kalman filter-based tracker for estimation. The Kalman tracker has higher motion estimation accuracy, as expected, since the Kalman filter greatly improves the quality of intraframe matching. However, the computational requirements are significantly greater since several matrices must be multiplied and inverted per frame.

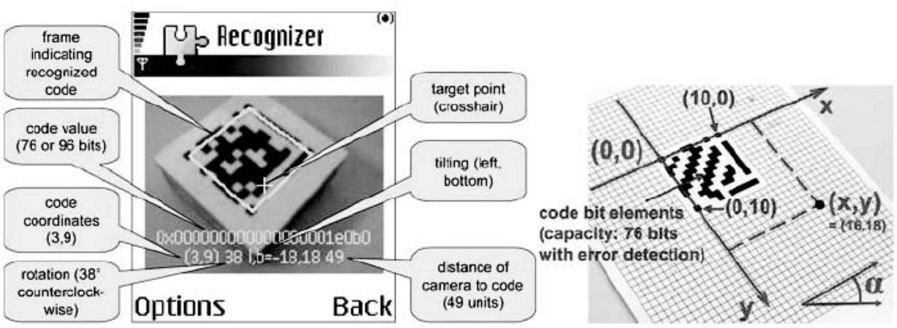
- Drab and Artner (2005) present a computationally inexpensive tracking system; however, their system has potential problems with repeating textures, and requires scenes with high dynamic range.

Tagging-Based Systems

Tagging-based tracking has also been researched by several groups. The main principle of these techniques is as follows: first, visual markers, printable with a standard printer, are created and placed in the environment. Then, these markers are detected in the captured video, and the 2-D and 3-D position of each visible marker (relative to camera position) and its rotation (relative to default orientation of the marker) are extracted. Additional information, such as an identification number, can also be detected with these solutions. Each proposed tagging technology has its own advantages and disadvantages:

- CyberCode (Rekimoto & Ayatsuka, 2000) is a visual tagging system based on 2-dimensional barcodes that can be recognized by CMOS and CCD cameras. CyberCodes encode 24 bits of data. In addition to the ID, the system can also compute the 3-D position of the tagged objects. Proposed applications for CyberCodes are augmented reality systems, various direct manipulation techniques involving physical objects, and indoor guidance systems.
- Rohs (2004) provide the Visual Codes system with an address space of 76 bits (83 bits without error detection) and a second guide bar, which allows the recognition of codes at a greater amount of tilting. The algorithm provides the relative x, y, and rotational motion of the phone, representing three degrees of freedom (DOF) input.
- Intelcom (http://www.intelcom.ru) has developed a software development kit for Nokia 7650/3650 for decoding data matrix codes. An example application generates SMS messages from the phone number and text stored in the code.
- Augmented reality and 3-D interaction research on mobile-camera-based tracking systems includes that of Möhring et al. (Möhring, Lessig, & Bimber, 2004), who track a color-coded 3-D marker to estimate 3-D camera pose, after an initial calibration step; and Hachet et al. (Hatchet, Pouderoux, & Guitton, 2005) who use a color-coded target in front of the camera to infer the 3-D DOF position.

Figure 2. Visual code parameters (left) and code coordinate system (right) (Rohs & Zweifel, 2005), with kind permission of Springer Science and Business Media



- The popular ARToolkit (Kato & Billinghurst, 1999) also provides visual markers. The toolkit provides optical tracking tools for detecting markers in a live video stream, extracting the 3-D position of the marker (relative to camera position) and its rotation (relative to default orientation of the marker).

The marker-based solutions suffer from the fact that a number of markers should be visible in a frame to detect the motion and orientation of the camera. Furthermore, the required presence of markers in the user's environment limits the range of interaction scenarios.

MAPPING CAMERA MOTION TO USER INTERACTION

As described, there are numerous proposed solutions computer vision technologies for enabling camera-based interfaces, and each solution has its advantages and disadvantages. For designing camera-based interfaces, the use of vision techniques have at least three variations:

- **Camera can be used as a pointing device:** The 1-D, 2-D, and 3-D position and movement of the camera can be acquired by tracking, and these data can used for various pointing tasks.
- **Camera can be used for gestural interaction:** The camera movement can be used as a low-level primitive for high-level gestural interaction techniques. For example, the user can interact with the applications by a set of gestures (e.g., by tilting and shaking the phone).
- **Camera can be used for interacting with the user's environment:** For example, instead of displaying the phone menu on the display of the device, the camera can be used to overlay the user interface in 3-D onto the video of the user's real world.

Next, we will discuss each variation and issues related to their use for interaction.

Using Camera as a Pointing Device

Direct-manipulation interfaces are particularly attractive for mass-market mobile interfaces because users can avoid learning commands and menu sequences, reduce their chance of errors, and keep their attention on the mobile device's small display. However, direct-manipulation interfaces are still difficult to realize on today's devices; currently, only key-modal interfaces are supported through joypad/direction keys and numerical keyboard. Stylus-based interaction is the most popular alternative, but this requires two-handed interaction, and has been shown to cause additional attentional overhead in users.

Using the camera as the pointing device on a mobile device is useful for many direct-manipulation interaction tasks. The range of tasks and their application create a rich set of design alternatives. Camera-based interaction allows for the following pointing tasks (Foley, Wallace, & Chan, 1984):

- **Selection:** Users can choose from a set of items by moving and tilting their phone, instead of pressing the phone's direction keys. This technique can be used for navigating through phone menus, quickly browsing through contacts, or selecting an image in the media library, for example.
- **Positioning:** Users can choose a point in 1-D, 2-D, or 3-D dimensions by moving their phone. Example uses of the position are controlling the mouse pointer on device's display, or selecting a hyperlink in a Web page.
- **Orientation/direction:** Users can choose a direction 1-D, 2-D, or 3-D space. The direction can be used for direction of a motion (e.g., to drag a file from one folder to another), to scroll a document vertically or horizontally, move an input message or a file to trash, or to rotate an image on the display, for example.

- **Path:** Users can rapidly perform a series of positioning and orientation operations. The path may be realized as drawing a curve in a drawing program, simulating handwriting by moving the phone as a stylus, or other direct manipulation tasks.
- **Quantification:** Users can specify a numeric value. The quantification can be used as one-dimensional or two-dimensional selection of integer or real values as parameters, such as the continuous zooming level, while viewing a Web page or a document.
- **Text:** Users can enter, move, or edit text in a two-dimensional space. The pointing device allows one to indicate the location of insertion, modification, or change. Other text-editing tasks, such as formatting a paragraph, can also be realized by moving the camera.

It is ultimately possible to implement a complete direct manipulation interface using the camera as the pointing device, but a way of switching the mode of the input for each of these tasks needs to be available to the user.

Using Camera as a Primitive for Gestural Interaction

In addition to the above low-level pointing tasks, camera can be used for gestural interaction. Gestures have traditionally been defined as a particular movement in front of the camera. However, in mobile interaction, gestures are generally defined by the motion of the camera instead. Each gesture can be defined as a different motion path, which is the output of tracking (Rohs & Zweifel, 2005), or can also include other high-level camera motions, such as shaking of the camera (Haro et al., 2005).

- For shake detection, Haro et al. (2005) determine the magnitude of the physical movement; they use motion history images (MHI) (Davis & Bobick, 1997), which were originally used for performing action and gesture recognition. An additional tracking algorithm provides four directions as application-level events, similar to mouse movement: up, down, left, and right, in the camera plane. The magnitude is also passed as an event where two states are possible: motion magnitude increasing or decreasing.

Figure 3. Possible combination of interaction primitives to build complex gestures for interaction (Rohs & Zweifel, 2005), with kind permission of Springer Science and Business Media

Combination	Interaction cue	Combination	Interaction cue
pointing & rotation	+ highlighted area	rotation & stay	
pointing & tilting	+ highlighted area	rotation & keystroke	
pointing & distance	+ highlighted area	tilting & distance	
pointing & stay	+ highlighted area	tilting & stay	
pointing & keystroke	+ highlighted area	tilting & keystroke	
rotation & tilting		distance & stay	
rotation & distance		distance & keystroke	

- Rohs and Zweifel (2005) propose and evaluate a number of physical gestures that form a basic vocabulary for interaction when using mobile phones. The proposed techniques are based on a visual code system that provides a number of orientation parameters, such as target pointing, rotation, tilting, distance, and relative movement. Their proposed framework defines a set of fundamental physical gestures that form a basic vocabulary for describing interaction when using mobile phones capable of reading visual codes. These interaction primitives can be combined to create more complex and expressive interactions.

Using Camera for Ubiquitous Computing and Augmented Realities

In addition to the tasks mentioned previously, camera input will play a different and more significant role in future mobile interfaces, together with the use of emerging ubiquitous computing and augmented reality paradigms. The interest in ubiquitous (or pervasive) computing has surged in the past few years, thanks to improving mobile processor and sensor technologies. The main characteristic of ubiquitous computing is to break away from the desktop interaction paradigm and move the computational power to the environment surrounding the user. To support this, ubiquitous computing requires user input to move beyond the textual input of keypad and selection from pointing devices, to perceptual interfaces that interact with users and their surroundings.

Augmented reality has emerged as one of the complementary fields that will be the mode of interaction for applications that combine mobile user interfaces with real-world interaction. AR provides a way to overlay computer-generated information (e.g., UI widgets or information) on top of real-world images. Although early AR work has focused on the use of see-through head-mounted displays, recent work has addressed handheld augmented reality, where the images are shown on a mobile device's display:

- Feiner et al. have presented one of the pioneering works in mobile AR interfaces. Their research on their MARS System (Mobile Augmented Reality Systems) began in 1996, and is aimed at exploring AR user interfaces, software, and application scenarios. They have proposed a set of reusable user interface components for mobile augmented reality applications (Feiner, MacIntyre, Höllerer, & Webster, 1997).
- The area of augmented reality on smart phones is still very new, but already there are several toolkit prototypes. The most well-known augmented-reality framework, the AR-toolkit (Kato & Billinghurst, 1999) has been used in a large number of augmented-reality research projects. Work is ongoing by several research groups to provide an efficient implementation of the toolkit available on cell phones and PDAs.
- One of the largest extensions to the AR-toolkit is Studierstube (Schmalstieg, Fuhrmann, Hesina, Szalavari, Encarnação, Gervautz, & Purgathofer, 2002), a framework that aims to make it easier for developers to create collaborative AR applications.

In addition to the frameworks described, a number of companies and developers are creating their own for mixed-reality applications, primarily in gaming. A large number of mixed-reality games for smart phones exist at the moment, ranging from camera-movement tracking to move on-screen crosshairs to simple body part region tracking. The most successful mixed-reality framework and platform for gaming at the moment is Sony's Eyetoy, which consists of a 60fps camera that is attached to a Sony Playstation 2. The processor analyzes incoming video to estimate the player's motion and to segment them from their environment. Eyetoy allows for games to perform game-control input recognition of: moving arms to certain locations to "touch" objects, moving arms up and down, moving the body side to side, and moving the body towards or away from the camera.

In augmented reality, accurate registration between the real and virtual data is essential. In practice, the calibration parameters recovered describing the camera's physical properties are good enough for rendering synthetic objects in augmented-reality applications. In the smart phone domain, camera calibration is significantly more difficult due to the large amounts of lens distortion present. Calibration techniques such as Zhang's accurate error-function minimization approach (1999) have been used successfully.

CAMERA-BASED UI PROTOTYPE

Towards the goal of building interactive applications, we have created a camera-based interaction prototype for smart phones. The prototype is similar in functionality to those on desktops, thus making it easier for computer vision and augmented-reality experts to work on smart phones more easily. Naturally, with smart phones there are hardware restrictions, so not all algorithms are presently possible. However, with a good framework it will be straightforward to extend the library with other algorithms, as well as other layers for content creators who are not vision experts.

While developing this prototype, we have faced a number of design issues. The split of the framework into low-level and high-level components was necessary to allow for expert and nonexpert developers alike to use the library to perform advanced camera input image and video processing. The low-level interface consists of the basic image and linear algebra operations themselves. Necessary geometric image operations, such as rotations, color conversions, image filtering as well as image statistics calculations, are provided. The low level also contains a set of linear algebra classes, including matrices, vectors, and functions such as singular value decomposition (SVD). The high-level interface is more focused on user interaction tasks. For instance, if a developer would like to determine whether the camera is being shaken, they can use library calls in the high-level interface to detect shaking, although the shake detection algorithm will be comprised of numerous calls to functions in the low level.

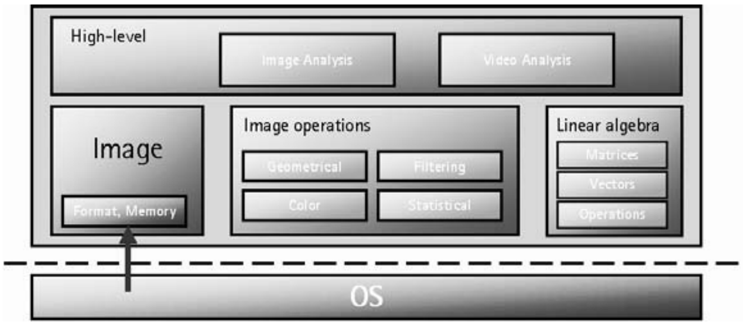
As described earlier, various tracking algorithms can be used for estimating the movement of the camera. In this chapter, we present a markerless, template-matching-based solution for tracking.

High-Level Algorithm Description

The tracking system was implemented on the Symbian OS. (See Haro et al., 2005 for details.) The process diagram of the tracker is presented in Figure 5. Our tracker uses the current and previous frame captured by the camera for tracking:

- First, the algorithm detects "corner-like" features in the new frame that are matched with the features found in the prior frame. Traditionally in computer vision, features include edges and corners. Edges are usually not significantly temporally coherent (i.e., they might change from one frame to another

Figure 4. Smart phone computer vision framework. This framework is the first step towards building camera-based applications built on a shared framework.



drastically), and corner features are too computationally expensive to find at many image locations while retaining real-time performance. We detect edges on both frames using the well-known Sobel filter. We use a threshold of 50 on both derivatives for each pixel, as this value results in a good number of feature candidates for typical scenes. Feature matching is performed between frames using template matching with 15x15 search windows, which we empirically found to be sufficient for our test hardware.

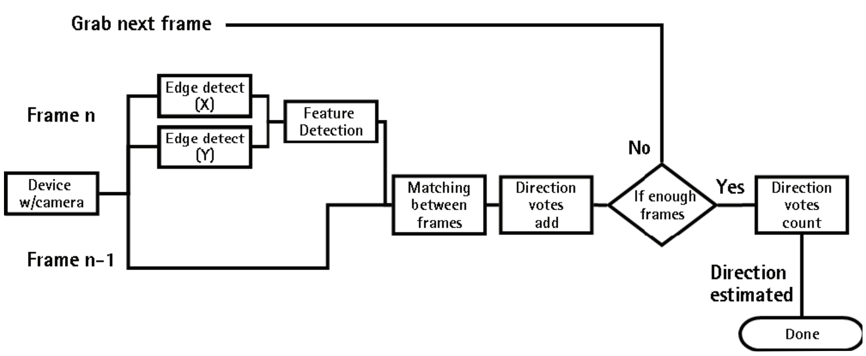
- Direction estimates are accumulated for a number of frames before a movement direction estimate is made. Direction voting is performed using variables, and the final decision on motion estimation is performed every four frames. This allows several frames to "vote" on the motion, keeping the scrolling from being incorrect due to any errors in other parts of the system.
- The directions of dominant camera motion are computed using the tracking algorithm, but their magnitudes are not known accurately. Camera motion magnitude must be calculated accurately to determine how to adjust the scroll speed in applications that need zoom control. We use motion history images (MHI) (Davis & Bobick, 1997) to estimate camera motion magnitude. Motion histories are encoded in single images such that a single image can be used for simple, robust, and computationally inexpensive gesture recognition. An MHI is computed by performing background subtraction between the current and previous frames. At locations where the pixel values change, the MHI is updated by decrementing by a predefined constant amount. By averaging the intensity values of the MHI, the average camera-motion magnitude is estimated.

Applications

We have implemented several test applications using the proposed prototype to clarify its strengths and limitations. We implemented the tracking algorithm and applications in C++ using the Series 60 second edition feature pack 2 SDK. Our test platform was a Nokia 6630 mobile phone that features an ARM 9 220mhz processor, 10 megabytes of RAM, 176x208 screen resolution, and a 1.3 megapixel camera capable of capturing frames at 15 fps.

Document viewer. Scrolling a document is a commonly difficult task on mobile devices. For instance, Web content designed for desktop computers is vertically much longer since mobile devices have narrower screens. In addition, joystick scrolling is especially difficult when scrolling vertically and horizontally. An alternative is to add an extra hardware button for scrolling. However, an extra button is not a preferable solution for mobile device manufacturers due to the lack of extra physical space on the device, along

Figure 5. Tracking algorithm

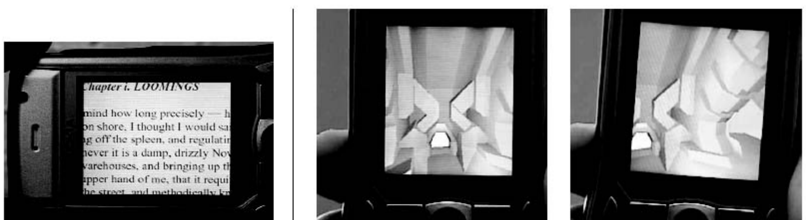


with additional manufacturing costs. In the document viewer prototype application that we implemented (Figure 6(a)), the user can vertically scroll documents by moving the device. Our approach is similar to *AutoZoom* and *GestureZoom* interfaces proposed by Patel et al. (Patel, Marsden Jones & Jones 2004). The scroll speed depends on how fast the user moves the device, which is much more intuitive than changing scrolling speed depending on how long the user presses the joystick, or via menu options and settings. One issue we identified in this application is that at some point, the user has to move the device more than they can reach. For example, if the user is scrolling to the right, at some point they will reach the physical limit of their arm's motion. To address this problem, we use the joystick as a "carriage return" that scrolls the document to the beginning of the next line and allows the user to move their arm back to the left again. After a carriage return, all tracked motion, except movement to the right, is ignored.

Zooming photo browser. As cameras become more widespread on mobile phones and storage size increases, managing photos becomes a more difficult task for the user, as large amounts of information must be viewed with limited input modalities. Current typical photo viewer applications show photo thumbnails as lists, grids, or 3-D carousels. Since image selection and scrolling are done with the joystick, the amount of time a user needs to browse their images is directly related to the number of images that they are browsing. Our photo browser test application (Figure 7) shows thumbnails of the user's photos in a grid layout. The user can scroll in four directions (up, down, left, right, in the camera plane) by physically moving the mobile device. In this case, it is difficult to view all the images, as some zoom control is required when looking for a particular image. If the zoom level is not properly set, it is difficult for a user to select a particular image from the set, as the scrolling will be too fast. To address this problem, we used the adaptive zooming technique introduced by Igarashi and Hinckley (2000). Adaptive zooming, based on the magnitude of the user's physical movement, keeps the scroll speed virtually consistent, allowing the user to browse more thumbnails by only moving the device faster.

3-D Game interaction. Creating an immersive 3-D experience is difficult on mobile devices due to the limited display size. The most immersive experiences are typically created using a combination of large displays reducing peripheral vision as much as possible and/or virtual environment navigation tied to the user's physical motion. In our prototype (Figure 6(b)), we map the user's physical motion to the view-point to create the illusion of a window into a 3-D world. Our renderer loads standard Quake III™ or Quake III Arena™ maps. Textures, light maps, curved surfaces, and lighting calculations are disabled for performance. The rendering is done using the OpenGL ES implementation available in the latest Symbian OS-based Series 60 SDK. Precomputed vertex lighting and fixed point calculations are used to improve performance due to the lack of a floating-point unit on our test hardware. The renderer is able to

Figure 6. (a) Camera-based interaction in a document viewing application. (b) Mapping physical motion to viewing direction creates illusion of a window into an environment.



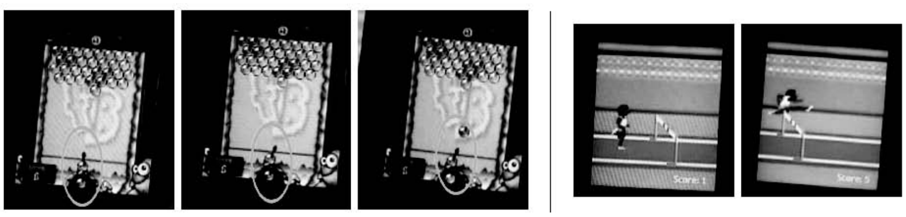
(a) (b)

Figure 7. Picture browser application; the application automatically adjusts the zoom level to help the user browse

realistically render lit virtual environments with several thousand polygons per scene at 3-10 frames per second, depending on the environment that is chosen. Navigation of the virtual environments is performed with a combination of physical motion and keypad presses. Actual movement in the environment is controlled by the keypad. The user looks around in the scene by physically moving the device around their body in the directions that they would like to look. We map the tracked camera motion directions to a trackball as in traditional mouse-based 3-D interaction. The combination of detailed environments, camera-based control, and interactive frame rate create a mobile user experience closer to that using additional hardware or larger displays.

2-D Game interaction. Camera-based user interaction can be used to enhance 2-D games as well as those that are 3-D. Camera motion can be used to add an additional element of interaction in games that require precise movements or very well-timed button presses. We created puzzle and action game prototypes to investigate these ideas using the camera motion and shake-detection algorithms presented. We modified the open source Series 60 port of the "Frozen Bubble" puzzle game (http://fbs60.sourceforge.net/), switching the game control from using the keypad to using the camera (Figure 8(a)). In our version, the user moves their device left and right to aim, and performs sudden shakes to launch their bubble. This has the effect of significantly changing gameplay, as careful arm motions are now required to aim, instead of a number of button presses, which increases the excitement as the game is now more physically based. We created a camera-based action game prototype as well. Using sprites and artwork from Konami's `Track and Field™' game for the Nintendo Entertainment System, a new game (Figure 8(b)) was created. A runner must jump over a never-ending number of approaching hurdles. To jump, the player must time the shaking of their device correctly so that the character does not crash into hurdles. Relying on the camera exclusively

Figure 8. (a) Players move the device left and right to aim, and shake the device to launch a bubble. (b) A jump command is issued by shaking the device at the correct time to avoid tripping on the hurdle.



(a) *(b)*

for input results in a game that is very simple to learn and understand but difficult to master, providing a new type of game. Shake detection is performed by thresholding the average intensity of the computed MHI.

DISCUSSION

In order to support intuitive and efficient user interaction, it is important to understand what kind of input is provided by the camera-based interfaces, what type of target task is considered, and what the current limitations are, given the output of the tracking algorithm. The success for a camera-based user interface is achieved when the following criteria are met:

- The speed and accuracy of the used computer vision technique should match the requirements and efficacy of the target task,
- The learning time should be minimal, particularly for gesture-based and augmented-reality interfaces,
- The computational cost and reliability of the used camera input processing techniques should match the target application and user's environment.

The most basic, but potentially most important input, that can be acquired from the tracking algorithm is the two-dimensional movement of the mobile device on a plane parallel to the camera in 3-D. With this type of data, the camera can be used as an input device to capture the device's movement in up/down, left/right directions, as well as its speed in each direction. Mobile camera-based input has restrictions, primarily due to limitations of mobile device hardware. Forward and backward motion cannot be detected with the current generation of mobile phones, so six degree–of-freedom movement is not supported yet. Forward/backward motion is possible to detect; however, this would increase computational demands and reduce the frame rate, impoverishing the user interaction.

Physical movement speed is another challenge for camera-based interaction. The algorithm must perform all of its video analysis in the time between camera frames being captured to support real-time interaction.

Thus, there are implicit limits on the computational complexity of the tracking. In addition, there is a fundamental assumption in tracking algorithms that each frame contains some portion of the prior frame. This assumption is motivated by the observation that users will typically not move their phones erratically when focused on a task. Users usually operate mobile phones with one hand. Mobile phones can also be used anywhere in an office, school, public, home, and so forth. Considering these environments, there are certain interactions that are not appropriate:

- **Precise tasks:** Precise motion is very difficult holding a mobile device with one hand. Interaction should not require operations like "move the device 2.5cm up," or "move the device 34 degrees from the horizontal line." As a result, camera-based interaction will probably be most useful when navigating large amounts of data, or zoom-level-dependent data.
- **Large motion:** This restriction is more serious in some environments, such as in crowded public locations. In such situations, it may be advantageous to provide a "clutch" to turn the tracking on/off. This would emulate the act of lifting a mouse once the edge of a desk is reached in traditional desktop interaction. In our informal testing, we did not provide a clutch; however, in commercial implementations this is a consideration to keep in mind.
- **Extended and/or frequent interaction:** Using single-handed operation, interactions that require extended time and/or frequent movement may fatigue users.

The camera-based input approach works best with coarse selections at different speeds and scales of data. It is critical that visual feedback follows physical motion and that the feedback differs according to motion speed, in order to provide an intuitive user experience. The most typical use case is moving the device to scroll UI content such as a list or a document.

CONCLUSION

In this chapter, we surveyed the technologies underlying mobile camera-based user interaction, and described the concepts and recent research results. We have presented a camera-based toolkit prototype, including design issues that we faced. We demonstrated our approach in several applications using 2-D and 3-D interaction. Initial results suggest that camera-based interaction has a great potential for future user interfaces.

While the proposed tracking solutions are computationally efficient and work well in practice in controlled environments, there are some situations that cannot be handled. Severe lighting differences will cause the template matching to stop working properly. Motion in front of the camera is ambiguous and can affect tracking results, as it is impossible to tell whether the camera is moving or not without other sensors. Shadows may confuse the tracking system, but there are known computer vision techniques for robust tracking in the presence of shadows that will be incorporated into the tracking algorithm once additional processing speed is available.

In the future, we would like to collect user feedback to determine how to improve user interaction further using mobile cameras, using qualitative and controlled experiments. While we applied the camera-based interaction to only viewpoint selections and simple gestures, we would like to investigate its application to more complex gesture-based and augmented-reality interfaces. In the future, we believe that a full camera-based mobile user interface is possible with the described approaches, potentially making the phone keypad unnecessary.

ACKNOWLEDGMENT

This work was partially supported by grants by Nokia Inc. and Bilkent University.

REFERENCES

Davis, J., & Bobick, A. (1997). The representation and recognition of human movement using temporal templates. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 928-934).

Drab, S., & Artner, N. (2005). Motion detection as interaction technique for games and applications on mobile devices. In *Extended Abstracts of PERVASIVE: Workshop on Pervasive Mobile Interaction Devices* (pp. 48-51).

Feiner, S., MacIntyre, B., Höllerer, T., & Webster, T. (1997, October 13-14). A touring machine: Prototyping 3-D mobile augmented reality systems for exploring the urban environment. In *Proceedings of the First IEEE International Symposium on Wearable Computers.*

Fishkin, K., Gujar A., Harrison, B., & Want, R. (2000, September). Embodied user interfaces for really direct manipulation. *Communications of the ACM*.

Foley, J. D., Wallace, V. L., & Chan, P. (1984). The human factors of computer graphics interaction techniques. *IEEE Computer Graphics and Applications*, *4*(11), 13-48.

Hachet, M., Pouderoux, J., & Guitton, P. (2005, April). A camera-based interface for interaction with mobile handheld computers. In *Proceedings of the 2005 Symposium on interactive 3-D Graphics and Games (Washington, District of Columbia. SI3-D '05)* (pp. 65-72).

Hannuksela, J., Sangi, P., & Heikkila, J. (2005). A vision-based approach for controlling user interfaces of mobile devices. *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*.

Haro, A., Mori, K., Capin, T., & Wilkinson S. (2005). Mobile camera-based user interaction. *Proceedings of ICCV-HCI 2005* (pp. 79-89).

Hinckley K., Pierce, J., Sinclair, M., & Horvitz, E. (2000). Sensing techniques for mobile interaction. *Proceedings of the 13th Annual ACM Symposium on User Interface Software and Technology (UIST 2000)* (pp. 91-100). New York, NY: ACM Press.

Igarashi, T., & Hinckley, K. (2000). Speed-dependent automatic zooming for browsing large documents. In *Proceedings of the ACM symposium on User interface software and technology (UIST 2000)* (pp. 139-148). New York, NY: ACM Press.

Kato, H., & Billinghurst, M. (1999) Marker tracking and HMD calibration for a video-based augmented reality conferencing system. In *Proceedings of the 2nd International Workshop on Augmented Reality* (IWAR 99), San Francisco.

Maes, P., Darrell, T., Blumberg, B., & Pentland A. (1995) The ALIVE system: Full-body interaction with autonomous agents. In *Proceedings Computer Animation* (pp. 11-18, 209).

Möhring, M., Lessig, C., & Bimber, O. (2004). Optical tracking and video see-through AR on consumer cell phones. In *Proceedings of Workshop on Virtual and Augmented Reality of the GI-Fachgruppe AR/VR*. (pp. 193-204).

Patel, D, Marsden, G., Jone,s S., & Jones, M. (2004). An evaluation of techniques for browsing photograph collections on small displays. In *Proceedings Mobile HCI* (pp. 132-143), Glasgow, Scotland.

Rekimoto, J., & Ayatsuka, Y. (2000). CyberCode: Designing augmented reality environments with visual tags. In *Proceedings of DARE 2000*.

Rohs, M. (2004). Real-world interaction with camera-phones. In *Proceedings of 2nd International Symposium on Ubiquitous Computing Systems (UCS 2004)* (pp. 39-48).

Rohs, M., & Zweifel, P. (2005). A conceptual framework for camera phone-based interaction techniques. In *Proceedings of Pervasive 2005*, (pp. 171-189). Schmalstieg, D., Fuhrmann, A., Hesina, G., Szalavari, Zs., Encarnação, L. M., Gervautz, M., & Purgathofer, W. (2002). The Studierstube augmented reality project. *PRESENCE - Teleoperators and Virtual Environments*, *11*(1).

Siio, I. (1998). Scroll display: Pointing device for palmtop computers. *Asia Pacific Computer Human Interaction*, 243-248.

Yee, K. P. (2003). Peephole displays: Pen interaction on spatially aware handheld computers. *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 1-8).

Zhang, Z. (1999). Flexible camera calibration by viewing a plane from unknown orientations. *International Conference on Computer Vision (ICCV'99)* (pp. 666-673).

KEY TERMS

3-D Interaction: A type of user interaction, where the output is displayed in 3-D (on stereoscopic or 2-D displays), and user input is received through 3-D interaction devices.

Augmented Reality: A field of computer research which deals with the combination of the real world with computer generated data.

Camera-Based User Interaction: A type of interaction, where the user interacts with an application by moving their device, and the captured camera video is used to estimate phone motion or interact with the real world.

Computer Vision: The analysis of image sequences, concerned with computer processing of images from the real world. Computer vision typically requires a combination of low-level image processing and high-level pattern recognition and image understanding to recognize important features in the image.

Gestural Interaction: A type of user interaction, where the user interacts with the computing device through a set of well-defined gestures. Gestures can originate from any bodily motion or state but commonly originate from the face or hand.

Ubiquitous Computing: A computing paradigm, where computation is integrated into the environment, rather than having computers as distinct objects. One of the goals of the field is to embed computation into the environment, and allow everyday objects to be used for interaction.

Zoom Control: Ability to control the zoom level in a user interface or an image or a document.

Section I
User Interface Design for Mobile Technologies

This section looks at many of the critical aspects concerned with effective design of mobile applications. The section begins with a series of chapters that discuss the adoption of ethnographic methods to inform the design of such technologies, including a selection of chapters that report on observed mobile device use and subsequent implications for design. This section covers issues such as how factors of user acceptance of mobile services can be used to guide the design of such technologies, as well as the impact of age and cognitive capacity on design. Chapters consider wearable technologies, the importance of contextual information in mobile application design, the design of in-car user interfaces, and issues surrounding the design and implementation of mobile learning applications. The section takes a look at adaptive and intelligent user interfaces for mobile computing, as well as tools for rapid prototyping, modeling, and simulation of mobile systems. The section concludes with a look to the future in terms of ecologies of interacting artifacts, reflecting an evolution from strictly mobile to more ubiquitous technologies.

Chapter XXXIII
3-D Visualization on Mobile Devices

Andrea Sanna
Politecnico di Torino, Italy

Fabrizio Lamberti
Politecnico di Torino, Italy

ABSTRACT

Recent improvements in technology of last generation mobile devices (smartphones, Personal Digital Assistants, Ultra-Mobile PCs) have opened up challenging new scenarios in 3-D ubiquitous visualization. Entertainment, cultural heritage, telemedicine, and distance learning are only a few of the disciplines that can take advantage of 3-D model visualization on mobile and hand-held devices. This chapter investigates the main methodologies used to display and navigate complex scenes and heavy datasets on mobile devices. Both local and remote rendering software techniques are considered with respect to solutions based on hardware acceleration. Moreover, issues related to the design of suitable graphics user interfaces are tackled. Finally, a solution for high performance visualization based on the remote rendering approach and enabling interactive manipulation of 3-D scenes composed by millions of polygons at 30 frames per second even on 50 Kbit/s wireless communication channels is presented.

INTRODUCTION

The advent of large-bandwidth wireless networks and improvements in graphics hardware have opened up exciting and intriguing new scenarios concerning ubiquitous visualization on mobile devices. Mobile phones and personal digital assistants (PDAs) are now able to display 3-D graphics contents and the demand for visualization applications is rapidly increasing.

In the past years, reduced hardware resources, limited bandwidth communication channels and inadequate graphics users interfaces (GUIs) strongly limited the use of mobile devices for 3-D graphics. Nowadays, mobile phones (often named smartphones), PDAs, and tablet PCs are able to connect to the Internet via high-speed networks (like for example IEEE 802.11 and UMTS) and high-performance CPUs, able to provide a support for multimedia applications, are becoming available. Several solutions integrating a CPU and a GPU (graphics processing unit) have also been developed (it is worth observing that a spread diffusion of graphics applications for desktop PCs began when accelerated graphics adapters appeared on the market). Moreover, hand-held devices can be equipped with a wide range of accessories allowing researchers to design and implement a complete and effective set of multimodal interfaces.

All these technological advances make ubiquitous graphics a new and challenging field of research. Today mobile devices are able to locally manage and display several thousands of polygons per second, as well as play high-resolution video streams in real time. Additionally, frameworks for managing mobile interactive visualization of highly complex datasets composed of millions of polygons/voxels have been developed. This means that ubiquitous 3-D visualization can now be performed in an efficient and effective way, thus helping scientists and doctors to achieve a clearer understanding of the nature of data under investigation, and supporting the entertainment market. Many fields of research and applications can take advantage of ubiquitous 3-D visualization: education, art, advertising, medicine, entertainment, and so on.

In order to provide an overall view on the state of the art of 3-D visualization on mobile devices, the chapter starts following a scheme based on the device workload. Thus, *local visualization* is considered first in the following Section. In this technique, the entire visualization process (geometry management, rendering, and display) is performed exploiting exclusively the computational resources of a mobile device often named fat client. An entire Section is then devoted to *remote visualization*. Here, a remote server is in charge of processing the geometry and rendering the scene. The mobile device, also called thin client, only has to display a set of still images or a video stream. Remarks concerning the design and the development of mobile GUIs for 3-D are then presented. Finally, the design and experimental results of a framework for high-performance 3-D visualization on mobile devices based on the remote rendering approach are presented.

LOCAL COMPUTATION

The term local computation refers to a situation where the entire geometry of the model can be stored on the mobile device and locally available computational resources are sufficient to render and display the scene (Zhou, 2006). The adjective sufficient can take different meanings according to the application fields being considered: a few frames per second can be considered sufficient when a limited interaction is required, but 30 frames per second, or more, can be necessary for video games or other “real-time” applications.

It is worth stressing that local computation requires data to be stored in the memory of the mobile device; beyond implications related to hardware resources, a problem of security can arise. Sometimes, it is preferred not to move visualization datasets outside “secure” environments, because data can contain reserved and/or highly confidential information. On the other hand, limited network resources are not involved in the visualization process, at least for single-user applications. Thus, visualization performances do not depend on network coverage and bandwidth.

From the software point of view, several solutions have been proposed to design and implement 3-D rendering engines for mobile devices. MiniGL (2006) and PocketGL (2006) were two of the first examples; they consist of a 3-D toolkit exporting a large number of utility functions needed to manipulate 3-D objects. They exhibit several similarities with OpenGL (Woo, Neider, Davis, & Shreiner, 1999). In fact, OpenGL code and tutorials can be used as a starting point to implement MiniGL- and PocketGL-based applications.

Another set of APIs named OpenGL ES, similar to OpenGL but specifically tailored for mobile devices, has been recently proposed (OpenGL ES, 2006). It is worth observing that the acronym ES means embedded system in order to outline the vocation of this library. Similarly to OpenGL, OpenGL ES is a set of low-level functions representing a subset of OpenGL calls capable of taking advantage of advanced embedded graphics. Several implementations are already available like Vincent (2006) and Hybrid's Gerbera (2006). Another open-source library similar to OpenGL is Klimt (2006), formerly known as SoftGL; Klimt is targeted to hardware independence and is available for many mobile platforms.

As Direct3D (Glidden, 1997) is a competitor of OpenGL, Direct3D Mobile (Microsoft, 2006) is likely to become a competitor of OpenGL ES. By using Direct3D Mobile, developers can take advantage of their existing desktop computer Direct3D skills and code to produce rich three-dimensional gaming environments for Windows Mobile devices. Direct3D Mobile takes full advantage of graphics hardware support, and enables the development of high-performance three-dimensional rendering on mobile devices.

Several sets of APIs for J2ME (Java 2 Platform, Micro Edition) have also been proposed. M3D (2006) is able to scale for both limited resource devices (like, for example, CPUs without support for floating point computation) and high-end devices equipped with specialized graphics hardware. M3G - JSR 184 (2006) is built as a wrapper on top of the Hybrid OpenGL ES API implementation; it is a J2ME optional package that allows moderate complexity 3-D graphics to be rendered at interactive frame rates on mobile devices.

Three-dimensional scenes can also be inspected on mobile devices by means of VRML (virtual reality modeling language) browsers. Pocket Cortona (Parallel Graphics, 2006) was one of the first VRML browsers for viewing VRML scenes on wireless devices; it is worth remarking that the latest version available can take advantage of the Xscale technology (Intel Xscale, 2006) to improve performances and reduce power consumption over a wide spectrum of hand-held and mobile devices.

Unfortunately, all the solutions mentioned are able to produce a smooth interactive visualization only for relatively simple scenes. For instance, Nokia 6630 smartphone is able to render about 60k triangles per second; if a frame rate of 30fps is required, only simple scenes modeled by a few thousands of polygons can be managed. Although the screen resolution of a hand-held device cannot always allow the user to appreciate highly detailed geometries, complex 3-D model visualizations can be required when displays provide a relatively large visualization area (as, for example, with last generation PDAs and existing Tablet PCs). In this case, the rendering engine has to be able to simplify the scene geometry to allow an interactive visualization.

Proposed techniques that modify the geometry (reducing the number of polygons without affecting in a noticeable way the appearance of the model) are based on level of detail (LOD) algorithms; the basic idea is to display a simplified version of the geometry when the scene is navigated, while the maximum level of detail is only displayed for fixed points of view. LOD algorithms can be classified as static or dynamic; a static LOD algorithm performs an off-line simplification of the geometry (the computational cost of simplification is not paid during the rendering) by producing a set of new models representing different approximations of the original objects. During visualization, the rendering engine chooses the appropriate representation for each object according to the distance from the observer (far objects are represented by less-detailed geometries). Static LOD algorithms are able to reduce computational costs; however, more memory space is needed to store the different representations. On the other hand, dynamic solutions are able to compute/simplify online the geometry of the scene. Dynamic strategies are able to produce smooth (almost unnoticeable) transitions between visualizations of different levels of detail, but they result in large computational costs often not affordable on mobile devices. Exhaustive surveys of generic multiresolution modeling techniques have been provided by Garland (1999) and Luebke (2001).

An example of LOD algorithm implemented for mobile devices has been developed by Zunino (Zunino, Lamberti, & Sanna, 2003). In this work, a continuous multiresolution modeling technique has been used to achieve a satisfactory trade-off between quality and interactivity. The proposed methodology allows the user to set a desired frame rate value that is used as an objective function. The scene complexity is scaled down when the current frame rate is lower than the threshold value, while it is increased if the frame rate trend is positive. Geometry decimations/refinements are performed according to a cost function able to take into account both the distance of the objects from the observer and the impact of topological changes on the visualization.

Another approach that can be used to reduce the computational cost on hand-held and mobile devices is based on image-based rendering (IBR) algorithms. IBR techniques use a limited set of key frames (considered as input images) to "interpolate" all the intermediate pictures (McMillan, 1999). An example of application of IBR algorithms for mobile devices has been proposed by Chang and Ger (2002). The designed framework consists of two parts: a model constructor and an interactive warper. The job of the model constructor is to convert 3-D models into depth images that are amicable to 3-D warping. The model constructor can be considered as a preprocessing step; therefore, it may run remotely on desktop computers rather than on mobile devices. The 3-D warper runs on mobile devices; it accepts user inputs and displays the new views interactively. An implementation of the 3-D warping algorithm presented above has been proposed by McMillan and Bishop (1995). Unfortunately, IBR techniques often produce artifacts due to the camera position; for this reason, they represent a feasible solution only in a limited number of visualization scenarios.

Figure 1 shows a typical local visualization scenario: the client stores in its local memory the 3-D model to be visualized (the scene can be downloaded from a remote server) and a local application (for instance, a VRML browser or a viewer based on OpenGL ES technology) uses CPU and GPU (if available) of the mobile device in order to process the graphics primitives and render the scene.

REMOTE COMPUTATION

Although local computation strategies can constitute a viable solution to visualize "relatively simple" scenes, geometries described by millions of textured polygons or large-volume datasets are beyond the capability of any mobile device.

A completely different approach for dealing with these geometries consists in dividing rendering and visualization workloads between a remote server and the mobile device. According to this philosophy, the only purpose of the mobile device is to display a flow of data (a sequence of still images or a video stream) coming from a remote server that is in charge of storing and managing the geometry, rendering the scene and sending resulting frames to the client (Zhou, 2006). The remote server can be a graphics workstation as well as a cluster of PCs or any other specialized hardware able to manage and render extremely complex models. It is worth remarking that in remote visualization strategies, only data representations are delivered to the mobile device that sends back user commands to inspect the scene, while data can be kept on the secure remote server. This solution

Figure 1. Local visualization

can require large bandwidth connections if high resolution streams have to be delivered; moreover, the client should provide a support for efficiently carrying out multimedia operations in order to perform on-the-fly processing of incoming video streams. A general classification of visualization algorithms based on a client-server approach has been provided by Martin (2000).

In remote computation, two different approaches are possible: computing a continuous set of still images or generating a video stream. A video, such as an MPEG stream, is able to take advantage of both compression of the single frame and temporal redundancy in a sequence of consecutive frames. On the other hand, the visualization of a video stream requires more computational resources than a single picture; this can be a problem for certain classes of hand-held and mobile devices not provided with powerful multimedia capabilities.

A simple but effective idea for addressing these issues was proposed by Stegmaier et al. (Stegmaier, Magallón, & Ertl, 2002); an X-server is remotely controlled by a PocketPC version of VNC - virtual network computing (Richardson, Stafford-Fraser, Wood, & Hopper, 1998). This approach is extremely attractive since it is application-independent and does not require any change in the interface. On the other hand, two main drawbacks can be identified. The transmission of the entire screen of the remote server can involve a large network load; hence, it could be inadequate for limited bandwidth channels. Furthermore, the advantage of being application-independent could constitute a drawback since the system cannot be optimized for any specific application.

In order to reduce network requirements, VirtualGL (2006) was proposed. The idea is to send a compressed picture to the mobile client instead of the raw screen of the remote console; different coding schemes have been evaluated by Stegmaier et al. (Stegmaier, Diepstraten, Weiler, & Ertl, 2003). Although a compressed frame requires less bandwidth than a raw one, computation resources needed at the client side to decode and display a compressed picture increase. Lamberti et al. (Lamberti, Zunino, Sanna, Fiume, & Maniezzo, 2003) showed that the frame rate obtained using a Compaq iPaq H3630 PDA is maximum when raw frames are sent to the client, since the mobile device is not able to efficiently decode and display compressed images.

The solution proposed by Lamberti et al. (2003) adopts a specialized remote server composed by a cluster of PCs, where the rendering is piloted by a distributed framework in order to take advantage of multiple hardware accelerated graphics adapters. An OpenGL application runs over each PC of the cluster, and the computation workload is split among all the available GPUs. A master PC is in charge of receiving contributions from the other units, reassembling the resulting frame and sending the picture to the mobile client. The authors demonstrated that this approach allows a visualization frame rate of 7 fps to be achieved (at a resolution of 120×120 pixels) that is independent of the scene complexity (in fact, the computational cost is paid at the remote server side).

A high performance grid, based on Web services, was also proposed by Grimstead et al. (Grimstead, Avis, & Walker, 2005) in order to allow mobile clients to display large medical datasets; the performance of this solution is greatly affected by the grid effectiveness, which actually depends on network connections. While a PC cluster is typically obtained by using high speed (Gigabit or fast-Gigabit Ethernet) switches, grid connections often suffer from high latency times, and are not always suited to real-time/interactive graphics applications.

A platform-independent framework written in Java was proposed by Engel et al. (Engel, Sommer, & Ertl, 2000); a similar solution, tailored for low-bandwidth systems, was presented by Beerman (2003).

Some commercial solutions have been also presented; Silicon Graphics developed Vizserver (Silicon Graphics, 2005), which is able to provide application-transparent remote access to high-end graphics resources for multiple simultaneous users. Similarly, IBM developed a product called Deep Computing Visualization (IBM, 2006).

Among the commercially available solutions, it is also worth recalling the recent RealityServer by

Figure 2. Remote rendering

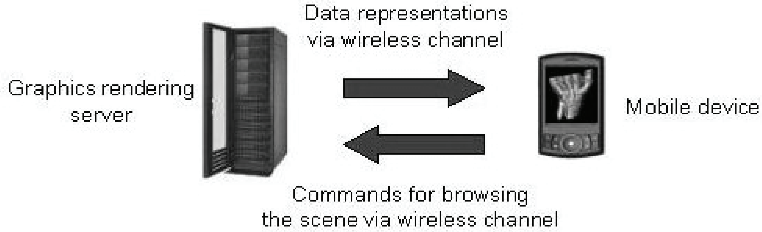
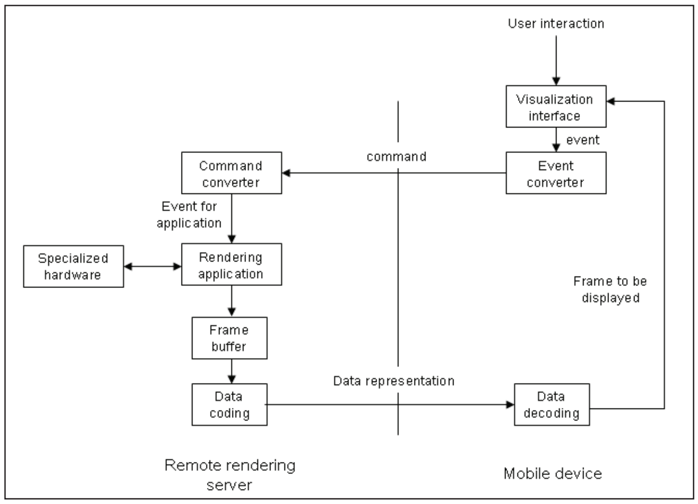


Figure 3. Sequence of operations in a remote rendering session



mental images (2006). RealityServer is a server-based platform allowing 3-D contents developed using CAD/CAM packages, to be navigated by remote client devices equipped with a Web browser that supports Flash or DHTML and JavaScript. RealityServer supports both local and remote rendering. When local rendering is used, a Web browser plug-in named RealityPlayer receives and manages the rendering of automatically-created reduced representations of the original 3-D content generated by the RealityServer. The system can additionally be deployed so that the server renders a photorealistic version of the original model, and sends it to the client for display after scene manipulation. When remote rendering is used, only a sequence or encoded stream of rendered images is delivered to front-end client devices. This allows in principle integration with any kind of client device. However, even if rendering and streaming can be processed in real time in certain rendering modes, bandwidth restrictions may adversely affect interactivity. In fact, according to *mental images*, white paper (2005), the visualization of a 3-D content on a smartphone/PDA at a resolution of 320×240 at 10 frames per second requires a constant bandwidth around 360 Kbit/s that largely exceeds even the average data rate of UMTS. Moreover, a 720 Kbit/s throughput is needed to handle scene manipulations at 20 frames per second.

Figure 2 shows a typical scenario where visualization on a mobile device is performed through a remote rendering technique. A remote rendering server computes the scene for the mobile client, and only a 2-D representation is sent over the network (as a flow of still images or as a video stream). The user application at the client side has to be designed to display data coming from the server and to capture commands issued to manipulate the scene. Commands have to be sent to the server to be processed; the rendering server receives commands from the client and updates the rendering of the scene accordingly. Figure 3 shows the sequence of operations of a typical remote rendering session in further detail.

GUI DESIGN REMARKS

The development of 3-D visualization environments on mobile devices requires not only performance constraints, but also user interaction considerations to be taken into account. In fact, 3-D navigation, and particularly 3-D manipulation, require complex GUIs to be developed. It is worth remarking that when a mobile visualization application only relies on local rendering, there is no need to completely replicate interfaces available in workstation-based scenarios. Thus, GUIs can be designed from scratch, taking into account the suggestions of research in the field on human-computer interaction (HCI). Nevertheless, when considering remote visualization solutions possibly involving collaborative sessions, different considerations arise. This is due to the fact that the remote server runs a visualization application characterized by its own GUI and interactions available at the workstation-based visualization site have to be made available also on a mobile device that is often characterized by a display limited both in size and resolution, a constrained keyboard, and a navigation button or stylus only partially emulating the behavior of a mouse. A communication channel is established between the remote server and the mobile device to report events generated by the user to the rendering server where they are translated into a modification of the visualization parameters. Below, some of the possibilities investigated in order to achieve effective visualization interaction will be reviewed. Their advantages and drawbacks will be also outlined.

In Grimstead et al.'s work (2005) complex 3-D datasets are represented through a scene graph hosting both polygons and voxels. While interacting with the dataset, the user can select an object by tapping on the display of the mobile device. A mapping algorithm allows the coordinates of the selection on the client device to be translated into a precise point in the remote scene graph. Depending on the specific manipulation context, it is possible to roll camera, move objects, and rotate around objects. The mobile client requests, from the rendering server, the set of available interactions for the particular object through a dedicated interface based on Web services. Available interactions are presented to the user through several menu options. A collaborative modality allows multiple users to concurrently manipulate a common scene. In this modality, each user is represented by an avatar linked to the camera, that is, to the user point of view. During a collaborative session, the user can follow the avatar of another user, thus sharing the same point of view. Additionally, depending on the security configuration being adopted, a user can control the camera associated with other users thus, making a sort of remote control of other devices possible. A possible application for this feature could be the use of a PDA to control a large-scale display. The authors expect to integrate more advanced collaboration functionalities, including the possibility of advertising points of view of common interest, the feasibility of linking to other cameras in order to implement guided tours within the virtual environment, and the adoption of audio and audio/gestural communications to improve cooperation. In the situation presented, a GUI is specifically designed for the application running on the mobile device. The only limitation to the complexity of the interface is related to the capabilities of the development environment available for the selected platform. The critical role is played by the mapping strategy that acts as a wrapper between the 2-D coordinate system of the mobile device display and the 3-D environment represented by the scene graph.

In Engel et al.'s work (Engel, Ertl, Hastreiter, Tomandl, & Eberhardt, 2000), a hybrid approach, based on both local and remote visualization, is adopted to support high-quality manipulation of volume data. A client application designed for mobile devices provides the user with a slicing tool deployed in Java/Java2D, permitting inspection of a complex medical dataset obtained from CT scans. Users can select a region of interest, and a low-quality 3-D rendering can be obtained on the device using Java3D. If a high-quality visualization is required, volume data can be transferred to a remote rendering server that actually performs the rendering, compresses the resulting frame, and transmits the 2-D picture ready for display to the mobile device. The local GUI allows the user to apply various transformations on the dataset including rotations, translations, and zooms. It is possible to change the transfer function, define orthogonal clipping planes, and perform iso-surface reconstruction. In this case, the GUI available at the remote-rendering server, based on Open Inventor, has been completely reconstructed on the mobile device using Java graphics functionalities. In this way, standard manipulators of Open Inventor are made available on the client side. Events generated by mouse, key presses, and widgets, including thumb wheels, sliders, push buttons, and context menus, are directly mapped onto the corresponding instances in the remote environment. The client has been tested on portable PCs, but it can be potentially migrated to Java-based hand-held devices.

Stegmaier et al. (2002) deployed VNC on a PocketPC device to remotely control a visualization application running on a rendering workstation. No particular solution is adopted to deal with the GUI on the mobile device. Thus, complex GUIs possibly available at the remote workstation are interleaved with visualization data and presented on the mobile device display. However, they are difficult to manage on the limited size display of a PDA device.

Woodward et al. (Woodward, Valli, Honkamaa, & Hakkarainen, 2002) presents a solution that enables engineers and other users to view and interact with 3-D CAD files using a wireless phone connection and a PDA device. It applies live video compression to transfer the continuously changing CAD view at the server workstation to the client PDA device. The client GUI lets the user rotate, pan, and zoom into the CAD model, as well as ask for distances, angles, and radii between points in the model. The author claims that other CAD interaction commands are straightforward to implement.

Burigat and Chittaro (2005) developed a dedicated application allowing mobile users to navigate VRML scenes by exploiting GPS-based localization information. Both local and remote rendering can be exploited to deal with scene visualization, depending on dataset complexity. In this case, no visualization console exists. There is no GUI to be replicated on the mobile device, and development efforts have been concentrated on the optimization of visualization performance. Thus, a simplified GUI has been designed to let the user perform manual or GPS-assisted navigation. The GUI provides a 3-D visualization area, status information, object details, and widgets to control the point of view when in manual mode.

Finally, a preconfigured solution is exploited by *mental images* (2006). In their RealityServer framework, imported 3-D contents are static, and existing graphics applications, possibly designed to manipulate the scene, cannot be reused. To allow interaction and customization, developers must write new application logic using a JavaScript-based API. The main drawback of this solution is that interaction logic is stored at the remote side together with 3-D data, and only limited customizations are possible at the mobile side.

HIGH-PERFORMANCE MOBILE 3-D REMOTE VISUALIZATION

Next, we present a complete framework for remote visualization in mobile environments based on distributed rendering and video streaming techniques. The proposed architecture can be transparently applied to professional visualization scenarios thus, enabling multiple mobile users to remotely control a variety of graphics applications based on OpenGL libraries. Experimental results showed that a 3-D scene, composed by more than three-million textured polygons can be visualized over wireless communication networks at about 30 fps at a resolution of 512 × 512 pixels on mobile and hand-held devices.

Remote Rendering Subsystem

The rendering subsystem at the server-side relies upon a software module named *distributed visualization service* (*DVS*) that is in charge of managing a remote rendering application deployed in a cluster-based distributed environment and communicating with a visualization application running on the mobile device. Through the *DVS*, multiple mobile users can control an OpenGL application running on a high-performance rendering facility during collaborative and interactive visualization sessions. It is worth observing that the *DVS* is also responsible for managing a scheduling subsystem permitting handling of concurrent manipulation commands that can be issued by both a visualization console at the server side and by simultaneous mobile users participating in the session.

The internal behavior of the *DVS* can be summarized as follows. Mobile clients connect to the *DVS* and start or join a visualization session by specifying several parameters including resolution, bitrate, and quality. Scene manipulation commands issued by mobile users are translated into a suitable format for transmission to the *DVS* over a wireless link. *DVS* processes incoming data, and extracts relevant information to be passed to the OpenGL callback layer for adjusting mapping and rendering configurations. Then, local graphics resources are used to execute OpenGL directives, and to generate the requested 2-D rendering frame to be sent to the mobile device for visualization. Graphics resources at the server side rely upon a Chromium (Humphreys, Houston, Ng, Frank, Ahern, Kirchner, & Klosowki, 2002) based architecture in order to distribute the execution of OpenGL rendering commands over a cluster of PCs. Chromium provides a software framework to unify the rendering power of a set of graphics accelerators distributed in cluster nodes through a virtualized interface based on OpenGL API. In a common Chromium architecture, a client submits OpenGL commands to one or more graphics servers. Each server is in charge of rendering only a part of the output image, which is treated as a tiled display and is later reassembled for visualization. In this way, even extremely complex scenes can be interactively rendered assuming that a sufficient number of graphics server nodes is available.

Chromium framework is based on a client-server paradigm relying on basic modules named *stream processing units* (*SPU*). Specific SPUs are responsible for subdividing 3-D context into tiles and for distributing graphics workload to server nodes (*Tilesort SPU*), carrying out the rendering (*Readback SPU*), sending back rendering parts to the client node (*Send SPU*) and finally, reassembling the overall image (*Render SPU*). In order to extend Chromium capabilities and enable it to support remote visualization, existing SPUs have been modified and new SPUs have been developed. In particular, a *Video SPU* has been developed to extract a frame-based video sequence from the assembled rendering framebuffer, and to serially distribute multiple copies of this sequence to ad-hoc components that are in charge of carrying out MPEG encoding and managing streaming to remote visualization clients according to the parameters passed during session initialization.

In particular, encoding is performed by a software module, called *Encoding Service*, that receives a frame-based raw video sequence and scales it to the requested resolution, converts rescaled frames to the YUV format suitable for MPEG processing, performs MPEG encoding generating an MPEG ES (Elementary Stream) bitstream (MPEG1, 1992), and finally passes encoded video to a streaming component. The *Encoding Service* module has been developed as a separate application that can be deployed on a dedicated machine communicating with the *Video SPU* over a TCP connection. In this way, the overhead on the client node is reduced and a performance speedup can be achieved. Beside that, multiple instances of the *Encoding Service* can be allocated (even on separate nodes), allowing for an effective management of heterogeneous client devices requesting video streams characterized by a different resolution, bitrate, and/or quality. The additional benefit of such a modular architecture is that the *Video SPU* and the *Encoding Service* can work in parallel. Thus, when a frame is being encoded, the *Video SPU* can carry out the next rendering cycle. According to this pipelined approach, when computing the frame rate that can be achieved in the video stream generation process, only the duration of the longest phase has to be taken into account.

MPEG frames generated by the *Encoding Service* are passed to the *Streaming Service* for distribution to mobile visualization devices using MPEG TS (Transport Stream). The MPEG TS format, defined in MPEG-2 specifications (MPEG2, 1994), has been chosen since it provides effective support for interactive transmission and visualization of multimedia contents over unreliable networks. Essential timing information is passed by means of time stamps inserted into each video data unit header. In particular, periodical transmissions of PCR (program clock reference) time stamps enable continuous synchronization of the receiver and transmitter clocks. PTS (presentation time stamp) indicates the time at which an encoded frame should be removed from the receiver buffer, instantaneously decoded and presented for display, while DTS (decoding time stamp) indicates the time at which an encoded frame should be instantaneously removed from the receiver buffer and decoded. Thus, streamed data incorporate sufficient information to carry out synchronization between encoder and decoder, allowing for the correct handling of out-of-order or delayed, and possibly corrupted, data delivery. In this way, data can be streamed directly over a UDP connection without introducing the overhead of additional protocols. Nevertheless, in the future, the effects of the introduction of alternative synchronization and control protocols (i.e., RTP, RTCP, RTSP) are expected to be evaluated.

The structure of the *Streaming Service* internal logic can be summarized as follows. The MPEG ES data flow received by the *Encoding Service* is first converted into an MPEG PES (Packetized Elementary Stream) (MPEG1, 1992) video sequence. This operation allows for the generation of variable length packets embedded with DTS and PTS information. MPEG PES stream is then converted into MPEG TS fixed-length data units including PCR information, which are placed in a FIFO output buffer for transmission. MPEG TS packets are extracted from the buffer according to the transmission scheme, and placed in a multicast UDP socket, thus, implementing simultaneous streaming to multiple mobile clients. The overall layout of the proposed architecture is illustrated in Figure 4.

Mobile Visualization Client

In the designed architecture, mobile devices equipped with a dedicated software application, called *Mobile Viewer*, simply display virtual 3-D scenes that are actually produced at a remote rendering facility. *Mobile Viewer* handles manipulation commands that are issued by mobile users through common input devices like touch screens, directional pads, and/or customizable buttons. *Mobile Viewer* application has been deployed using Gtk+ 1.2 and X11 on both an HP iPaq H5500 PDA running Familiar Linux 0.6.1 and on a Compaq T1000 Tablet PC running Linux RedHat 8.0. A porting of the source code also permitted evaluation of application performances in a Microsoft Windows XP-based native installation on the selected Tablet PC. The architecture of the client application can be summarized as follows (see Figure 4). A *Visualization GUI* component is in charge of handling events generated by the interface and passing them to an *Event Generator* module that converts them into a format suitable for transmission. The *Visualization GUI* module is also responsible for supervising the render area where the video stream generated by the *DVS* is actually displayed. An *MPEG Decoder* module receives the MPEG TS video stream distributed by the remote *Streaming Service*, it extracts synchronization data, decodes MPEG frames, and displays ready images according to received timing information. Video processing capabilities depend on a general purpose open-source multimedia library derived from the VLC player (Videolan, 2006). An example showing the usage of the proposed architecture is given in Figure 5.

Figure 4. Layout of proposed streaming-based remote rendering architecture

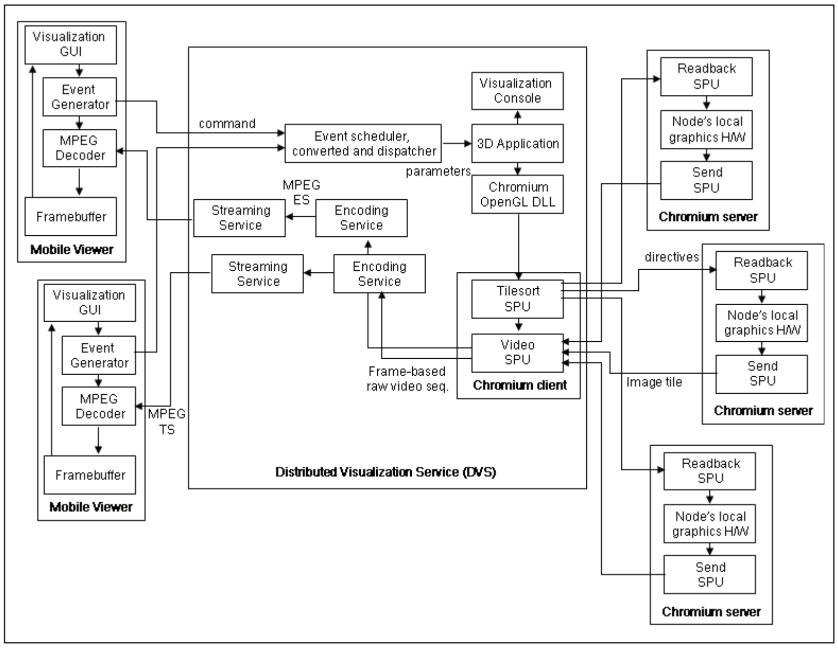
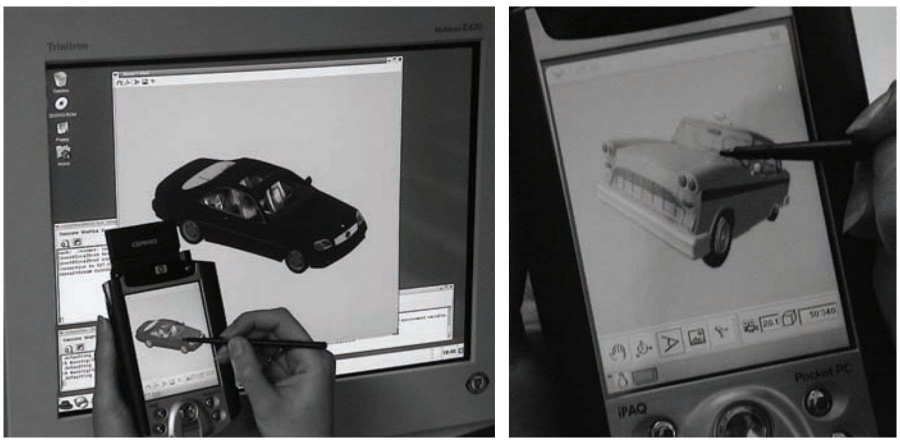


Figure 5. a) Remote visualization session involving a PDA and a visualization console located at the DVS site. b) Detail of user interaction with the remote 3-D scene.



(a) *(b)*

Experimental Results

The effectiveness of the proposed architecture has been evaluated in remote visualization scenarios where mobile users equipped with PDA and Tablet PC devices simultaneously interact with a general purpose 3-D surface rendering application running on an accelerated graphics back-end. The main aim is to prove that such a remote visualization framework allows for highly interactive collaborative sharing of realistic virtual worlds, between a high-end visualization console and multiple mobile users, regardless of the complexity of the scene being considered. Since visualization frame rate and overall latency experienced at the mobile client both constitute the main limitations of existing remote rendering architectures, we provided a quantitative analysis of the critical parameters, thus giving a precise measure of the achieved degree of interactivity. In particular, the time needed to render a single image (t_{render}), to generate the corresponding encoded frame (t_{MPEG}) and to display it on the mobile device ($t_{streaming}$) has been measured. Other important parameters include the time needed for a command to reach the *DVS* ($t_{command}$), the frame rate at the *DVS* (fps_{DVS}), the frame rate at the mobile device (fps_{client}), and the average bitrate ($bitrate_{client}$).

Experiments have been carried out on a chromium-based rendering subsystem including up to eight nodes, each running RedHat Linux Fedora 3 and equipped with a Pentium IV 2 GHz, an nVidia Quadro FX-1100 AGP graphics accelerator with 128-MB video memory, 256-MB main memory, and a GB Ethernet network card. We demonstated that, for a given visualization resolution, interactive scene manipulation can be experienced by simply removing the bottleneck given by the frame rate at the *DVS*, assuming that the available bandwidth is capable of supporting video streaming requirements. Measurements for a 240×240 rendering resolution and streaming video size and for increasing scene complexity and rendering nodes are illustrated in Figure 6. Corresponding frame rate at the *DVS* is reported in Figure 7. Since the *Mobile Viewer* is able to manage MPEG visualization at up to 30 fps, frame rate at the client side is not reported since it always corresponds to fps_{DVS} when fps_{DVS} is lower than 30 fps, and to approximately 30 frames per second when fps_{DVS} is larger than 30 fps. It can be observed that for high complexity scenes, t_{render}

Figure 6. Performance with a rendering resolution and streaming video size of 240×240 pixels for increasing number of polygons and video servers

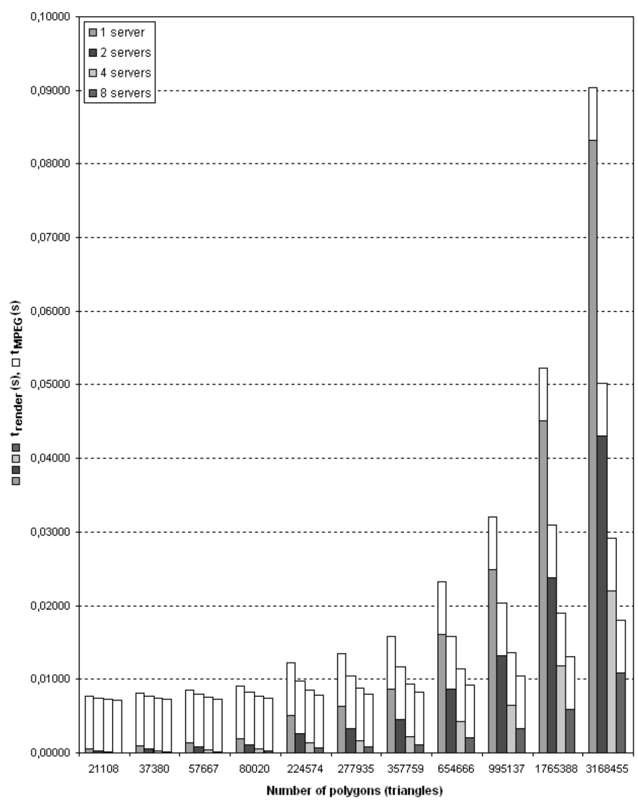


Figure 7. Frame rate with a rendering resolution and streaming video size of 240×240 pixels for increasing number of polygons and video servers.

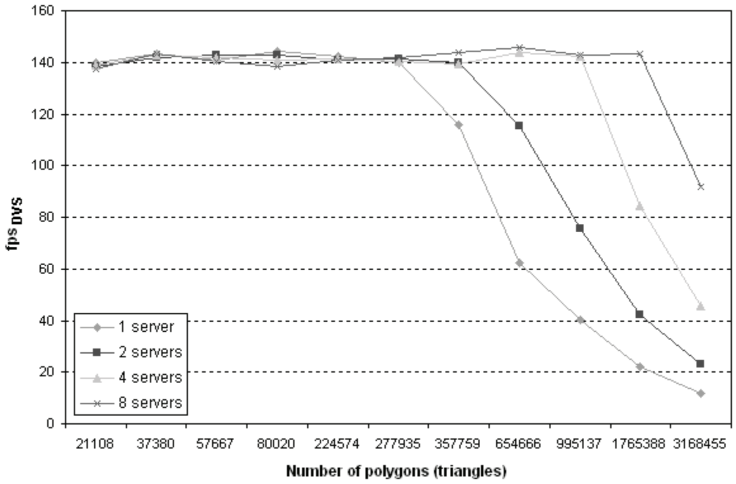
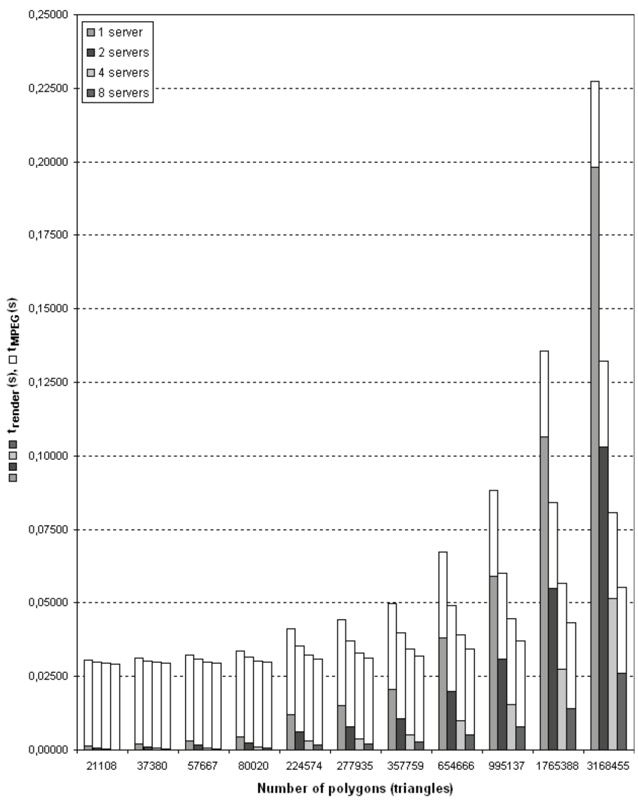


Figure 8. Performance with a rendering resolution and streaming video size of 512×512 pixels for increasing number of polygons and video servers.



becomes predominant over t_{MPEG} and the degree of interactivity is greatly reduced. Nevertheless, by increasing the number of rendering servers, t_{render} can be reduced, and interactive frame rates can be experienced. It is worth observing that all the considered scenes can be visualized at the maximum frame rate using only four servers. Results for a rendering resolution and streaming video size of 512×512 pixels are reported in Figures 8 and 9. In this case, without introducing a rendering distribution layer, even scenes with less than one million polygons can be navigated only with a low frame rate. However, using eight rendering servers, all the scenes can be visualized on the mobile device at 30 fps.

To conclude the evaluation of system behavior, several experiments were carried out to assess the impact of the main parameters of the *Video SPU* and *Encoding Service* as well as of the communication link being used on the overall latency between scene manipulation and display update. For this, four Chromium server nodes were used to allow the navigation of a high complexity 3-D scene (one million polygons per frame) at different rendering resolutions and visualization sizes ranging from 240×240 to 1024×1024 pixels. Moreover, different bitrates were used for the MPEG stream. Experiments were carried out in various wireless networking environments including GPRS, UMTS and WLAN IEEE 802.11b. Overall latency for all the considered configurations is reported in Figure 10, where the contributions of $t_{command}$, t_{render}, $t_{MPEG,}$ and $t_{streaming}$ and the corresponding frame rate are accurately quantified.

According to experimental results, we can conclude that latency in WLAN environments allows for very interactive scene manipulation, with an average response time of 0.42 s. Latency measured in GPRS- and UMTS-based communication scenarios can be possibly unacceptable when dealing with applications requiring real-time like interaction. Nevertheless, they can certainly ensure reasonable response times (around 0.83 s) in all those professional contexts where the goal is to enable visualization of highly complex 3-D scenes over large geographical distances.

CONCLUSION AND FUTURE TRENDS

In this chapter, a review of the main solutions for developing 3-D visualization applications on mobile devices is presented. Local computation-based techniques simply exploiting rendering resources available on the mobile side are considered first. Here, graphics support can be provided by general purpose CPUs or by specific GPUs. In the first case,

Figure 9. Frame rate with a rendering resolution and streaming video size of 512×512 pixels for increasing number of polygons and video servers.

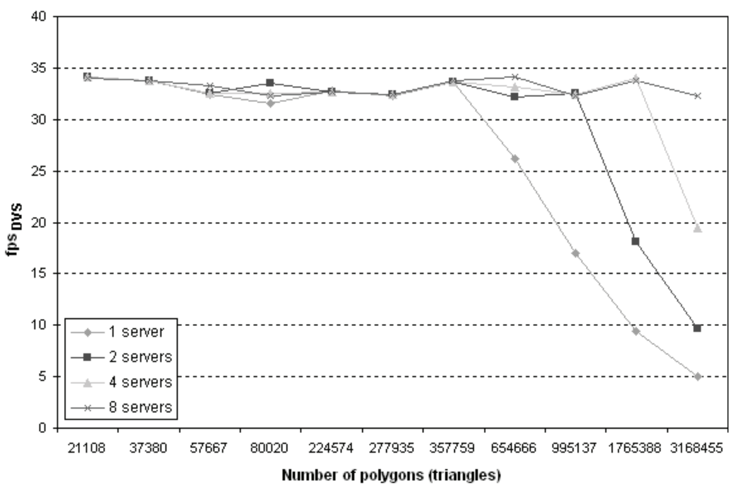
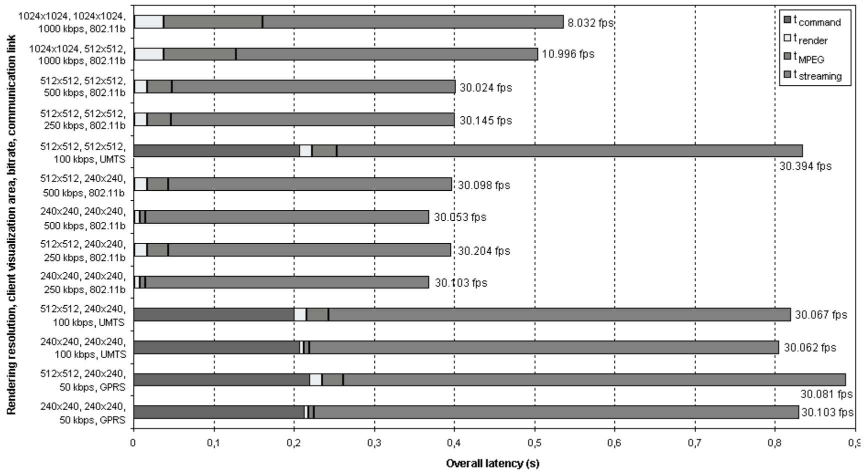


Figure 10. Impact of rendering resolution, visualization area size, video stream bitrate and communication channel on the degree of usability experienced by the mobile user in terms of overall latency and frame rate



only limited complexity scenes can be rendered at interactive frame rates. On the other hand, when specialized hardware is available, complex scenes can be smoothly navigated. Nevertheless, manipulation of extremely complex scenes required in specialized applications like remote diagnosis, collaborative CAD/CAM sessions, and scientific visualization becomes possible only through remote computation/visualization techniques. The application of remote visualization to the mobile world is relatively new, since obstacles to its diffusion have been overcome only in recent years, thanks to the advances in multimedia and communication technologies. Nevertheless, a solution allowing local-like manipulation of highly complex 3-D worlds on hand-held devices by means of distributed rendering and streaming techniques is presented in this chapter, and aspects related to the degree of interactivity perceived by mobile users are investigated in depth.

As regards future research activities in the field of mobile 3-D, it is worth observing that a crucial role will be certainly played by hand-held GUIs, 3-D APIs for mobile and embedded graphics accelerators. Next-generation GUIs will have to be able to provide effective interaction capabilities. In fact, one of the most limiting factors to the diffusion of mobile 3-D is related to the poor human-computer interfaces available on mobile terminals, including limited display and constrained input devices. Thus, it will have to be decided whether to present the mobile user with the same GUI available at the remote site, to recreate a simplified version of the interface or to design specific interfaces specifically tackling actual device constraints. From the software point of view, the future trend seems to be very similar to the scenario already seen for desktop PCs, where OpenGL and Direct3D represent the main technologies for developing 3-D graphics applications. Today, OpenGL ES and Direct3D Mobile are candidates for dominating the market of APIs for mobile systems. The trend for mobile hardware is also clearly outlined: all PDAs, smartphones, and other mobile devices will soon be equipped with specialized hardware for 2-D and 3-D graphics. Thus, local visualization will become increasingly attractive and feasible. However, remote visualization techniques, like the one presented in this chapter, will continue to play a key role in all those scenarios in which the complexity of the scene exceeds the hardware capabilities of the client device, and in which only data representations can be delivered over the network because of security constraints.

ACKNOWLEDGMENT

This chapter is part of a work developed within the European Research Project VICTORY, "Audio-VIsual ConTent search and retrieval in a distributed P2P repository." VICTORY is funded under the Information Society Technologies (IST) thematic program of the European Commission's Sixth Framework Program.

REFERENCES

Beerman, D. (2003, October). *CRUT: Event distribution and image delivery for cluster-based remote visualization*. Paper presented at IEEE Visualization 2003—Workshop on Parallel Vis. and Graphics, Seattle, Washington.

Burigat, S., & Chittaro, L. (2005). Location-aware visualization of VRML models in GPS-based mobile guides. In *Proceedings of the 10th International Conference on 3D Web Technology* (pp. 57-64).

Chang, C. F., & Ger, S. H. (2002). Enhancing 3D graphics on mobile devices by image-based rendering. In *Proceedings IEEE Third Pacific-Rim Conference on Multimedia* (pp. 1105-1111).

Engel, K., Ertl, T., Hastreiter, P., Tomandl, B., & Eberhardt, K. (2000). Combining local and remote visualization techniques for interactive volume rendering in medical applications. *IEEE Visualization*, 449-452.

Engel, K., Sommer, O., & Ertl, T. (2000). A framework for interactive hardware accelerated remote 3D-visualization. In *Proceedings of the Eurographics—IEEE TCVG Symposium on Visualization* (pp. 167–177).

Garland, M. (1999). Multiresolution modeling: Survey & future opportunities. In *Eurographics '99 State of the Art Reports* (pp. 111-131).

Glidden, R. (1997). *Graphics programing with Direct3D*. Addison Wesley Longman.

Grimstead, I. J., Avis, N. J., & Walker, D. W. (2005). Visualization across the pond: How a wireless PDA can collaborate with million-polygon datasets via 9,000km of Cable. In *Proceedings of the ACM/SIGGRAPH Web3D Symposium* (pp. 47-56).

Humphreys, G., Houston, M., Ng, R., Frank, R., Ahern, S., Kirchner, P., & Klosowki, J. T. (2002). Chromium: A stream-processing framework for interactive rendering on clusters. In *SIGGRAPH 02 Conference Proceedings* (pp. 693–702).

Hybrid's Gerbera Website. (2006). Retrieved June 12, 2006, from http://www.hybrid.fi/

IBM, Deep Computing Visualization Website. (2006). Retrieved June 9, 2006 http://www-03.ibm.com/servers/deepcomputing/

Intel Xscale technology specifications. (2006). Retrieved June 12, 2006, from http://www.intel.com/ design/intelxscale/

Klimt Website. (2006). Retrieved June 12, 2006, from http://studierstube.org/klimt/

Lamberti, F., Zunino, C., Sanna, A., Fiume, A., & Maniezzo, M. (2003). An accelerated remote graphics architecture for PDAs. In *Proceedings of the ACM/SIGGRAPH Web3D Symposium* (pp. 55-61).

Luebke, D.P. (2001). A developer's survey of polygonal simplification algorithms. *IEEE Computer Graphics and Applications*, *21*(3), 24-35

M3D Website. (2006). Retrieved June 12, 2006, from http://www.jcp.org/en/jsr/

M3G - JSR 184 Web site. (2006). Retrieved June 12, 2006, from http://www.hybrid.fi/ main/m3g/

Martin, I. M. (2000). *Adaptive rendering of 3D models over networks using multiple modalities* (Tech. Rep.). IBM T.J. Watson Research Center.

McMillan, L. (1999). Image-based rendering: A new interface between computer vision and computer graphics. SIGGRAPH Computer Graphics Newsletter—*Applications of Computer Vision to Computer Graphics*, *33*(4), 61-64.

McMillan, L., & Bishop, G. (1995). Plenoptic modeling: An image-based rendering system.

In *SIGGRAPH 95 Conference Proceedings* (pp. 39–46).

Mental images RealityServer Web site. (2006). Retrieved June 12, 2006, http://www.mental images.com/2_3_realityserver/

Mental images (2005). *RealityServer Functional Overview*. (Tech. Rep. No. 005-01). Retrieved June 5, 2006, from http://www.mentalimages.com/cgi-bin/rs-whitepaper.cgi

Microsoft Direct3D Mobile (2006). Retrieved January 2, 2007, from http://msdn.microsoft.com/library/default.asp?url=/library/en-us/wcemultimedia5/html/wce50 oriDirect3DMobile.asp

MiniGL Website. (2006). Retrieved June 12, 2006, from http://www.dsbox.com/minigl.html

MPEG1 - ISO/IEC 11172. (1992). *Coding of moving pictures and associated audio for digital storage media at up to about 1,5 Mbit/s.*

MPEG2 - ISO/IEC 13818. (1994). *Generic coding of moving pictures and associated audio information*.

OpenGL ES Web site. (2006). Retrieved June 12, 2006, from http://www.khronos.org/opengles/

Parallel Graphics Web site. (2006). Retrieved June 12, 2006, from http://www.parallelgraphics.com/

PocketGL Web site. (2006). Retrieved June 12, 2006, from http://www.pocketgear.com

Richardson, T., Stafford-Fraser, Q., Wood, K. R., & Hopper, A. (1998). Virtual network computing. *IEEE Internet Computing, 2*(1), 33–38.

Silicon Graphics, Inc. OpenGL Vizserver Web site (2005). Retrieved December 22, 2005, from http://www.sgi.com/software/vizserver/

Stegmaier, S., Diepstraten, J., Weiler, M., & Ertl, T. (2003). Widening the remote visualization bottleneck. In *Proceedings of the 3rd International Symp. on Image and Signal Proc. and Analysis* (pp. 174-179).

Stegmaier, S., Magallón, M., & Ertl, T. (2002). A generic solution for hardware-accelerated remote visualization. In *Proceedings of Eurographics—IEEE TCVG Symp. on Visualization* (pp. 87-94).

Videolan project Web site. (2006). Retrieved June 12, 2006, from http://www.videolan.org

Vincent Web site. (2006). Retrieved June 12, 2006, from http://ogl-es.sourceforge.net/

VirtuaGL project Web site (2006). Retrieved June 12, 2006, from http://virtualgl.sourceforge.net/

Woo, M., Neider, J., Davis, T., & Shreiner, D. (1999). *OpenGL programming guide: The official guide to learning OpenGL*—ver. 1.2 (3rd ed.). Addison-Wesley.

Woodward, C., Valli, S., Honkamaa, P., & Hakkarainen, M. (2002). Wireless 3D CAD viewing on a PDA device. In *Proceedings of the 2nd Asian Mobile Computing Conference* (pp. 83-87).

Zhou, H. (2006). *A survey on ubiquitous graphics* (Tech. Rep.). Hong Kong University of Science and Technology. Retrieved December 18, 2006, from http://www.cse.ust.hk/ ~zhouhong/publications.html/

Zunino, C., Lamberti, F., & Sanna, A. (2003). A 3-D Multiresolution rendering engine for PDA devices, In *SCI 2003 Proceedings, vol. V* (pp. 538-542).

KEY TERMS

API: An application programming interface (API) is a language and message format used by an application program to communicate with the operating system or some other control program such as a database management system or communications protocol.

GPU: Graphics processing units are specialized logic chips devoted to rendering 2-D or 3-D images. Display adapters contain one or more GPUs for fast graphics rendering.

GUI: A graphics user interface (GUI) is a particular case of user interface for interacting with a computer that employs graphic images and widgets in addition to text to represent the information and actions available to the user. Usually, actions are performed through direct manipulation of the graphic elements.

Local Visualization: In local visualization (or local computation) graphics tasks are carried out by exploiting the capabilities of the actual device.

Mobile/Portable/Hand-Held Devices: PDA, Tablet PC, Smartphone: Hand-held devices are pocket-sized computing devices, typically using a small display for user output and a miniaturized keyboard for user input. Input and output devices can be combined into a single touch-screen based interface. These devices are becoming increasingly popular amongst those users requiring the assistance and convenience of a conventional computer in a mobile and wireless environment.

Multimodal Interfaces: Humans perceive the world through senses; communication through one sense is known as a mode. Multimodal interfaces try to combine two (or more) modes of communication to enhance user interaction.

Polygon: A polygon is a simple closed curve consisting entirely of a finite number of line segments. A polygon mesh is a 3-D object comprising a certain number of polygons joined at their edges and vertices.

Remote Visualization: In remote visualization (or remote computation) heavy graphics tasks are delegated to a high-end graphics server that actually performs the 3-D rendering and generates a 2-D frame that can be visualized on a remote device (possibly characterized by limited hardware resources).

Rendering: Creating an image representing objects designed in a three-dimensional modeling program.

Streaming Video: A one-way video transmission over a data network. It is widely used on the Web as well as on private intranets to deliver video-on-demand. Unlike movie files that are played after they are downloaded, streaming video is played within a few seconds of requesting it, and the data is not stored permanently in the computer.

3-D Computer Graphics: 3-D computer graphics are works of graphic art that are created with the aid of digital computers and specialized 3-D software.

Voxel: A voxel (volume pixel) represents a quantity of 3-D data just as a pixel represents a point or a cluster of points in 2-D data. It is used in scientific and medical applications that process 3-D images.

Chapter XXXIV
Navigation Support for Exploring Starfield Displays on Personal Digital Assistants

Thorsten Büring
University of Konstanz, Germany

ABSTRACT

Due to advances in hardware technologies, mobile devices are increasingly capable of handling large-scale data sets. While this development broadens the application scope of smartphones and PDAs, it also means that high information loads must be displayed on very limited screen real estate. A solution to this problem may be provided by starfield displays. Starfield displays maximize the data-pixel ratio by presenting data inside a zoomable 2-D scatterplot. However, a drawback is that once users have zoomed into the information space, they tend to become lost, due to the clipping of orientation cues. The chapter summarizes the research results of recent projects that were conducted to improve the navigation and orientation features of starfield displays on small screens. Several approaches, such as smooth zooming, overview window, and fisheye view, have been implemented and user-tested. The results may support interface designers when targeting mobile devices.

INTRODUCTION

The continuous advances in mobile hardware technology and the ubiquitous availability of wireless networks lead to novel application domains for smartphones and personal digital assistants (PDAs). While these devices were originally developed for simple personal information management (PIM) tasks, they may soon serve as a truly mobile alternative to notebooks in many business scenarios. Major software companies, such as SAP and Oracle, are already responding to this development by providing initial extensions for mobile devices. Equipped with a PDA, field sales staff can access and browse the company database to retrieve product information while being on the road.

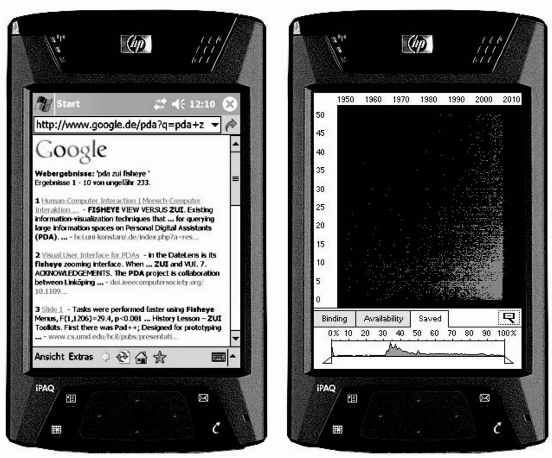
A bottleneck that impedes the development of mobile applications, in particular for data retrieval, is the small screen size of the devices. Current mobile interfaces tend to rely on conventional list- or table-based representations that make inefficient use of screen space. Having retrieved a large number of search results, users are forced to scroll and flip pages to identify records they are interested in. Even for comparably small data sets as in the left view in Figure 1, this approach is tedious, slow, and error-prone.

A more promising interface strategy for small screens is that of starfield displays (Ahlberg & Shneiderman, 1994). Starfield displays are interactive 2-D scatterplots that rely on space-preserving encodings rather than on textual representation of data. In the desktop world, these interfaces have been found to improve the performance when searching in movie databases (Ahlberg & Shneiderman, 1994) and housing databases (Williamson & Shneiderman, 1992), or to aid drug discovery in commercial retrieval frameworks such as Spotfire (Ahlberg, 1996).

As shown in the right view in Figure 1, in a starfield display, each data object is visualized by a small symbol that is spatially mapped against two scatterplot axes of data attribute dimensions. In this way, many thousands of items can be displayed on a single PDA screen, while users are provided with an effective overview that clearly reveals clusters, trends, and statistical outliers (Tufte, 1983). The problem with scatterplots, however, is that visual clutter from overlapping items is difficult to avoid. To prune visual clutter, starfield displays allow the users to explore the information space by moving between multiple representations and viewpoints (Dix & Ellis, 1998). This interaction is based on the metaphor of a zoomable user interface (ZUI). ZUIs follow the assumption that navigation in information spaces is best supported by tapping into our natural spatial and geographical ways of thinking (Perlin & Fox, 1993). In ZUIs, data objects are organized in space and scale. Users navigate this space by performing zooming (changing the scale) and panning (movement at constant scale) operations.

Developing effective starfield displays for pen-driven mobile devices is a challenging task. In this chapter, we investigate two major aspects of the design. One is how to provide appropriate input mechanisms for controlling zoom and pan operations. On PDAs, primary input commands are limited to screen taps and a small set of hardware buttons. This places severe constraints on the interaction design. The other major design aspect deals with orientation inside the zoomable scatterplot. A general drawback with ZUIs is that, due to the clipping of orientation cues, users may eas-

Figure 1. Comparing the data-pixel ratio of a list-based search engine output and a starfield display on a PDA



ily loose their orientation once they have zoomed into the information space (Card, Mackinlay, & Shneiderman, 1999) (Hornbaek, Bederson, & Plaisant, 2002). This problem becomes increasingly troublesome as less display real estate is available, and is therefore particularly significant for mobile devices. To reduce orientation problems, the literature proposes three solutions (Spence, 2001): smooth zooming (B. Bederson & Meyer, 1998; Shneiderman, 1996; Shneiderman & Plaisant, 2005), overview+detail (Plaisant, Carr, & Shneiderman, 1995), and focus+context interfaces (Spence & Apperley, 1982). We implemented and adjusted these approaches to improve the navigation and orientation in starfield displays on PDAs. This chapter introduces each interface developed, and summarizes the results of the user evaluations that were conducted to compare the solutions in terms of their usability. For an in-depth discussion of the evaluations see (Büring, Gerken, & Reiterer, 2006a; Büring & Reiterer, 2005; Büring, Gerken, & Reiterer, 2006b). The chapter concludes with an outlook on how to further improve starfield displays by incorporating visual cues for off-screen target locations and supporting advanced scaling interaction, such as speed-dependent automatic zooming.

SMOOTH-ZOOMING

The first project is a proof-of-concept demonstration of a smooth-zooming starfield display for PDAs. Despite the potential benefits for visualizing data on small screens, there have been only a few attempts to port interactive scatterplots to mobile devices. The first was the PalmMovieFinder (Dunlop & Davidson, 2000), which is based on the ideas of Ahlberg and Shneiderman's FilmFinder (Ahlberg & Shneiderman, 1994) and provides access to a collection of 71 movies. Exploration is limited to a simple two-level zoom in which a quarter of the display is magnified. More recent starfield applications are (Burigat and Chittaro (2005) and (Dunlop, Morrison, McCallum, Ptaskinski, Risbey, & Stewart, 2004), which focus on map-based visualizations. Again, only discrete zoom steps are supported.

Interface

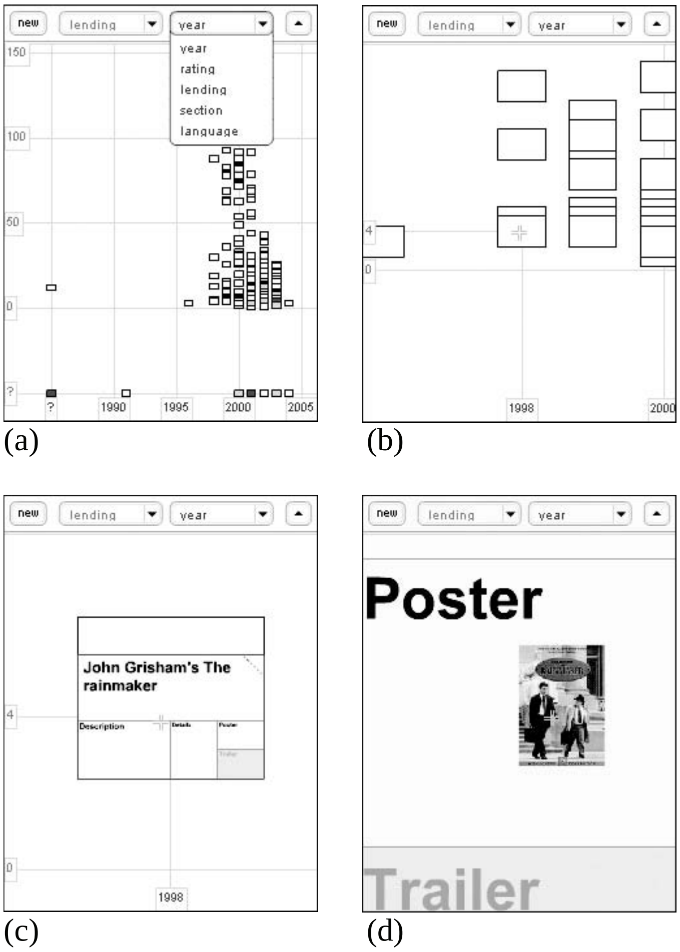
Our smooth-zooming prototype is implemented in Macromedia Flash and visualizes a movie database of 335 items. Having submitted a keyword query, retrieval results are represented by small rectangles inside a scatterplot (see Figure 2a). By default, the diagram dimensions are mapped to year of release (X) and lending frequency (Y). Other variables can be selected from the two color-coded combo boxes at the top of the interface. Alternative mappings are popularity rating, section, and language. Upon selecting a new mapping from a combo box, the scatterplot distribution and the axes labels are rapidly updated.

The prototype supports semantic and geometric zoom techniques to overcome the general drawbacks of scatterplot visualization, and to provide a work-around for those problems that apply specifically to scatterplot visualizations on small screens. Geometric zoom means that when zooming in on an object of interest (e.g., a street corner), the relevant map clipping simply becomes magnified. In contrast, an object that is zoomed semantically changes its visual representation depending on the amount of real estate available to it (Perlin & Fox, 1993).

As shown in the starfield display in Figure 1, large information spaces may cause data representations to become very small and thus hard to select accurately. Moreover, many data sets have an uneven distribution, causing the items to overlap and cluster together. Zooming reduces the information density and enables users to view a small clipping of the diagram surface in detail. To support orientation during the view shift, smooth scaling transitions have been proposed to help users to preserve their sense of position and context (Shneiderman, 1996).

Zooming in the prototype separates steering and scaling into two distinct actions. First, the users denote their current focus by tapping the screen with the stylus. A blue cross-hair appears. To zoom in on that position, users press and hold the upper button of the four-way PDA rocker switch. A smooth zoom animation is triggered, during which the rectangle items are continuously magnified

Figure 2. Smooth-zooming prototype: (a) default zoom-level, (b) geometric-zoom, (c) semantic-zoom, (d) subcontent-unit



and the view gradually shifts to center the focus point. The benefit of zooming in the scatterplot is that, at a sufficiently high magnification level, items can be accurately selected, even if they are overlapped to a great extent by other items (see Figure 2 b). Users can stop the animation at any time by releasing the button. Zooming out to the default scale is achieved by pressing and holding the lower rocker-switch button. To pan the view, users simply drag the diagram with the stylus.

While a scatterplot visualization is an effective tool for generating an overview of a large data set on limited screen real estate, it fails to provide detail information. In desktop applications, such information is often presented on-demand in a separate frame, but on mobile devices, the screen real estate is too limited to reserve permanent space for a widget that is only temporarily of interest to the users. Another solution would be to use pop-ups, but this technique drags the user attention away from the diagram. On closing pop-ups, users need time to reorientate inside the scatterplot. To provide a smooth and intuitive shift from overview to detail information, the prototype once more employs the zooming feature. At a certain scale level, the interface switches from geometric to semantic zoom. This causes the rectangles to turn into record cards, that is, into frames that display the content of the DVD item the rectangle represents (Figure 2c). Apart from the title, a record card also holds multiple subsections with additional information such as movie details, posters, and trailers. To view the data, users tap the corresponding subsection and perform a zoom-in operation to magnify the section's content to a readable size (see Figure 2d). Overall, the semantic zoom allows users to rapidly switch between overview and detail information without leaving the scatterplot visualization. Hence, orientation disruptions are avoided.

Evaluation Abstract

An informal user test based on observation and interviews was conducted with six participants (2 male and 4 female) in a laboratory setting to receive some general feedback on the intuitiveness of the starfield visualization and interaction. For the user test, the starfield application was installed on a Hewlett-Packard iPAQ hx4700 Pocket PC with Windows Mobile 2003 and Flash Player 6. The device featured a 624 MHz processor, 64 MB SDRAM, and a 64K color VGA touchscreen. During the experiment, participants were first asked to try out the interface on their own while thinking aloud. Subsequently, they had to solve a set of 12 retrieval tasks.

During the test the participants quickly understood the basic principles of the application and thus, were able to solve the test tasks without requiring much support. In particular, the concept of using semantic zoom to switch from abstract representation to a content view was much appreciated and proved to be as intuitive as expected. On the other hand, users found it increasingly difficult to orientate with a growing magnification of the scatterplot. Accordingly, we observed frequent zoom-out operations to the default scale. Another problem revealed by the user test concerns the zoom interaction. The sequential procedure of first setting a focus point and then pressing a button to manipulate the zoom level was found slow and tedious. Users comments indicate a clear preference for a one-step and purely pen-based zoom interaction.

OVERVIES+DETAIL

While smooth zooming supported users in the informal user test and was rated as a positive user experience, it could not prevent users from getting lost in the information space. This result also matches previous research (Card et al., 1999; Hornbæk et al., 2002). A more powerful solution to preserve user orientation may be provided by overview+detail (o+d) interfaces. O+d interfaces are characterized by a multiwindow layout, where one window is used to present details while the other one gives an overview of the information space (Plaisant et al., 1995). In addition to supporting orientation, overviews can also offer additional interaction possibilities such as panning and scaling. A difficult design question when dealing with o+d interfaces on small screens relates to the layout of the two windows. For both views, the usability increases with a larger size, since the presentation becomes less cluttered. However, a larger detail view means a smaller overview, and vice versa. This trade-off problem assumes that the interface uses the most common side-by-side layout of the two windows. Alternative layouts include transparent or on-demand overviews that overlap with the detail view. These techniques require less screen real estate, but typically make the interface more difficult to understand and use (Plaisant et al., 1995; Shneiderman & Plaisant, 2005).

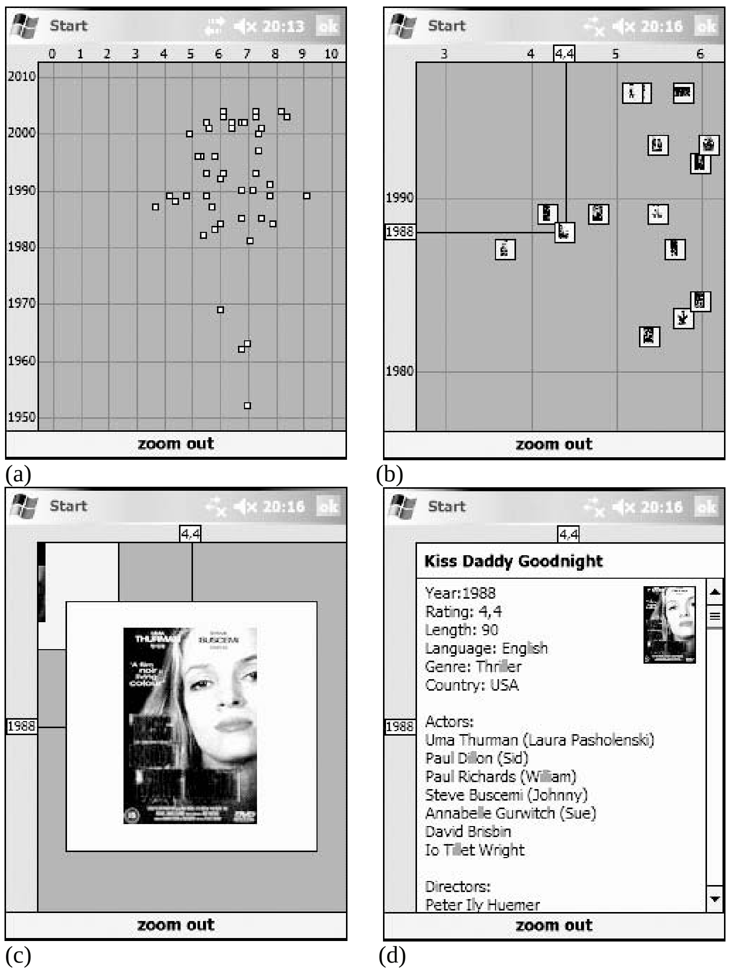
Mobile interfaces make increasing use of overview windows to support orientation. One example is the recent Nokia Web browser for S60 3rd ed. devices. When the software detects extensive scrolling, it displays a miniature image of the Web page. Despite the growing acceptance of o+d interfaces on small screens, there has not been much research on that topic as yet. A few studies have been conducted in the context of desktop computers. The results give the impression that users typically prefer o+d interfaces, while user performance varies with different task types and application domains. (Hornbæk & Frøkjær, 2001), for instance, investigated the usability of a linear, a fisheye, and an o+d interface for electronic documents. They found that all but one user preferred the o+d interface, and that essays written with it received higher grades from experienced tutors. On the other hand, users completed reading and question-answering tasks fastest when using the fisheye view. Similar results were found in a study on Web browsers (Baudisch, Lee, & Hanna, 2004): the participants preferred the o+d interface, but the overall performance was best when using a fisheye distortion. Another study specifically researched the usability of zoomable user interfaces with and without overviews for map navigation (Hornbæk et al., 2002). Again, the participants preferred the o+d interface, but when the participants used a map with a semantic structure (multiple levels with different content granularity), the authors found that the additional window significantly decreased the user performance. It was hypothesized that the rich navigation cues of the semantic zoom render an overview rather unnecessary.

Interfaces

To research whether o+d interfaces can be exploited to improve navigation in starfield displays, we developed two interfaces: one that featured an overview (overview interface) and one that relied solely on the detail view (detail-only interface). Both interfaces were implemented using the Microsoft .NET compact framework 1.1.

The detail-only interface (Figure 3a) shows a simplified layout of the original Flash prototype. Database items are encoded as small rectangles and positioned inside a scatterplot diagram. The

Figure 3. Detail-only interface: (a) default zoom-level, (b and c) zooming in, (d) record card



axes are set to the variables of popularity rating (X) and year of release (Y). Tapping the scatterplot causes the rectangle closest to the pen position to be highlighted. This strategy also ensures that even very small representations can be accurately selected. Highlighting an item moves it to the foreground and the user is provided with the movie's exact attribute values by displaying additional labels on the scatterplot axes.

In line with the user suggestions from the informal usability test, we modified the zoom feature to be fully controllable via pen events. While touching the screen highlights an item, holding the pen for more 150 milliseconds will trigger a fluent zoom-in operation on that item. Hence, users can highlight (set a focus point) and magnify items in a single action. It is also important to notice that, unlike in the Flash prototype, users are now prevented from zooming into the empty space between objects. A zoom operation is always bound to an item.

During the zoom, the highlighted item is magnified and centered. The information space contains three distinct semantic levels that are displayed depending on the current scale. In the default scale, the rectangles are too small to display any content information (Figure 3a). On magnifying them, the second semantic level is triggered. The items are enhanced with movie posters that scale together with the outlining rectangle (Figures 3b, c). Once the item is scaled to fill the entire window space, the system switches to the third semantic level. In a quick animation, the poster zooms out to a smaller size and is positioned at the upper right corner of a scroll pane. The free screen space is used to display the textual content of the movie data (Figure 3d).

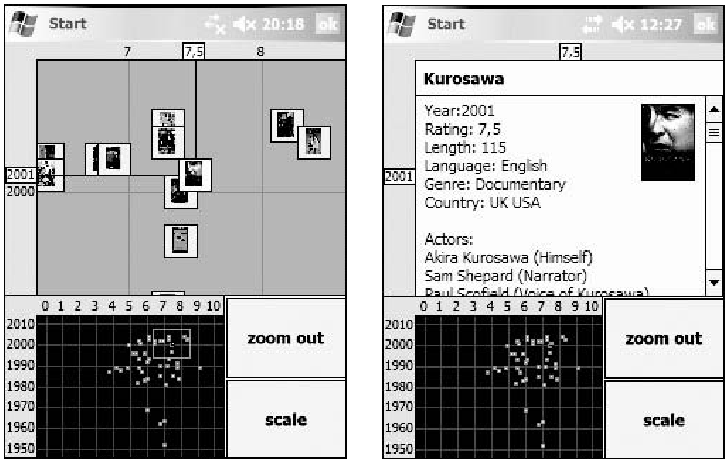
Users can interrupt a zoom operation at any time by lifting the pen. Moving from the default scale to the content view of the scroll pane magnifies the information space 40 times and takes 1.8 seconds. When zooming out by pressing the button at the bottom of the interface, the item grows smaller and changes its representation back to a poster and then to a simple rectangle. If not previously interrupted, the operation automatically stops at the default scale. Panning is again implemented as the action of dragging the information space with the stylus.

The **o+d interface** features a multiwindow layout, where the upper view is similar to the one in the detail-only application, but its height is decreased by about 35%. The free space is used to display an additional overview window (Figure 4). The sizes of the two windows have been chosen such that the axes units used on the overview can be the same as those on the detail view, with the labels remaining legible.

The overview window presents a miniature view of the entire information space at a static scale. Each dot on the overview is a rectangle on the detail view. To support the users' orientation, the overview includes a field-of-view box represented by a yellow rectangle. The rectangle shows the portion of the information space that is currently displayed on the detail view. During zooming, users are thus provided with off-screen context that is missing in the detail-only application. Detail view and overview are tightly coupled, which means that manipulating one view immediately updates the other one. If users highlight an item on the detail view, the corresponding dot on the overview turns red. Panning on the detail view shifts the field-of-view box.

Apart from visualizing off-screen content, the overview also provides additional navigation functionality. An alternative way to pan is to drag the field-of-view box on the overview. The detail view is continuously updated. Users may also jump to different locations by tapping outside the field-of-view box on the overview. The yellow rectangle automatically moves to center the pen position. To zoom on the overview, users first press the toggle button labeled *scale*, and then draw a new field-of-view box by dragging the pen (bounding-box zoom). To remain consistent with the given ratio of the views, the box is limited to a fixed aspect ratio. Zooming on the overview is limited to the first two semantic levels of the information space. Hence to retrieve the textual content of an item, users still have to use the detail view.

Figure 4. O+d interface to improve orientation



Evaluation Abstract

To compare the usability of the two approaches, we conducted an experiment in which 24 participants had to complete 12 search tasks on an 85-item movie collection using both PDA starfield applications. The study was run on a Hewlett-Packard iPAQ hx4700 Pocket PC with Windows Mobile 2003. For a full report of the experiment, see (Büring et al., 2006a).

Analyzing the results of the user test, we found that participants were significantly faster when using a detail-only ZUI compared to an overview-supported ZUI (379.34 seconds compared to 452.65 seconds, $F(1,23)=16.5$, $p<0.001$). This result is quite surprising, since the logged interaction data shows that, as expected, the overview window had reduced the need for long-distance panning and zooming. We hypothesize that the loss in performance is due to the time required to frequently switch between the two views in the o+d interface. This corresponds to previous research (Baudisch, Good, Bellotti, & Schraedley, 2002). Moreover, the effect may have been amplified by the additional cognitive load introduced by the increased interface complexity. Another explanation would be that the rich navigation cues provided by scatterplot labels reduced the benefit of having an overview window (Hornbaek et al., 2002).

Unlike in (Hornbaek et al., 2002), our experiment did not show a significant difference in preference between the two ZUIs (10 participants preferred the overview interface, 13 the detail-only interface, one subject was unsure, X^2 $(1,N=23)=0.391, p<0.532$). Participants who voted for the detail-only interface preferred the larger size of the detail window and the simplicity of the single-view layout. While they appreciated the orientation support, some users reported problems with zooming and panning on the overview. Due to the small size of the window, they found these functions rather imprecise and difficult to use. A larger overview would probably have improved the interaction but, as discussed earlier, could only be provided at the expense of a smaller detail view.

FISHEYE

While the o+d interface could not improve task-completion times or user satisfaction, the previous experiment, nevertheless, showed that overview information can reduce the need for unnecessary navigation.

A way to exploit this benefit without forcing users to switch between multiple views is to integrate both focus and context in a single view. A popular implementation of such an approach is the fisheye view (Furnas, 1986), a distortion concept that shows the focus in detail and the remote regions in progressively less detail. Other well-known focus+context solutions on desktop computers include the bifocal display (Spence & Apperley, 1982), the perspective wall (Mackinlay, Robertson, & Card, 1991), and hyperbolic views (Lamping, Rao, & Pirolli, 1995). Regarding mobile devices, an early example of a PDA application that utilizes a fisheye view is the WEST Web browser (Björk, Holmquist, Redström, Bretan, Danielsson, Karlgren, 1999). WEST uses a proxy server to split a Web page into several chunks that are then presented as small cards in a left-to-right and top-to-bottom fashion on a PDA. There is always one card in focus that is displayed at a larger size than the other cards. The users can move the focus to enlarge different content chunks, or they can zoom into a focused card to view it at full size. Later examples of mobile fisheye views are DateLens (Bederson, Clamage, Czerwinski, & Robertson, 2004) and AppLens (Karlson, Bederson, & SanGiovanni, 2005). These applications visualize calendar and application data in a tabular interface. The users can enlarge or minimize cells by tapping them with the stylus and thus, distribute display size according to their focus.

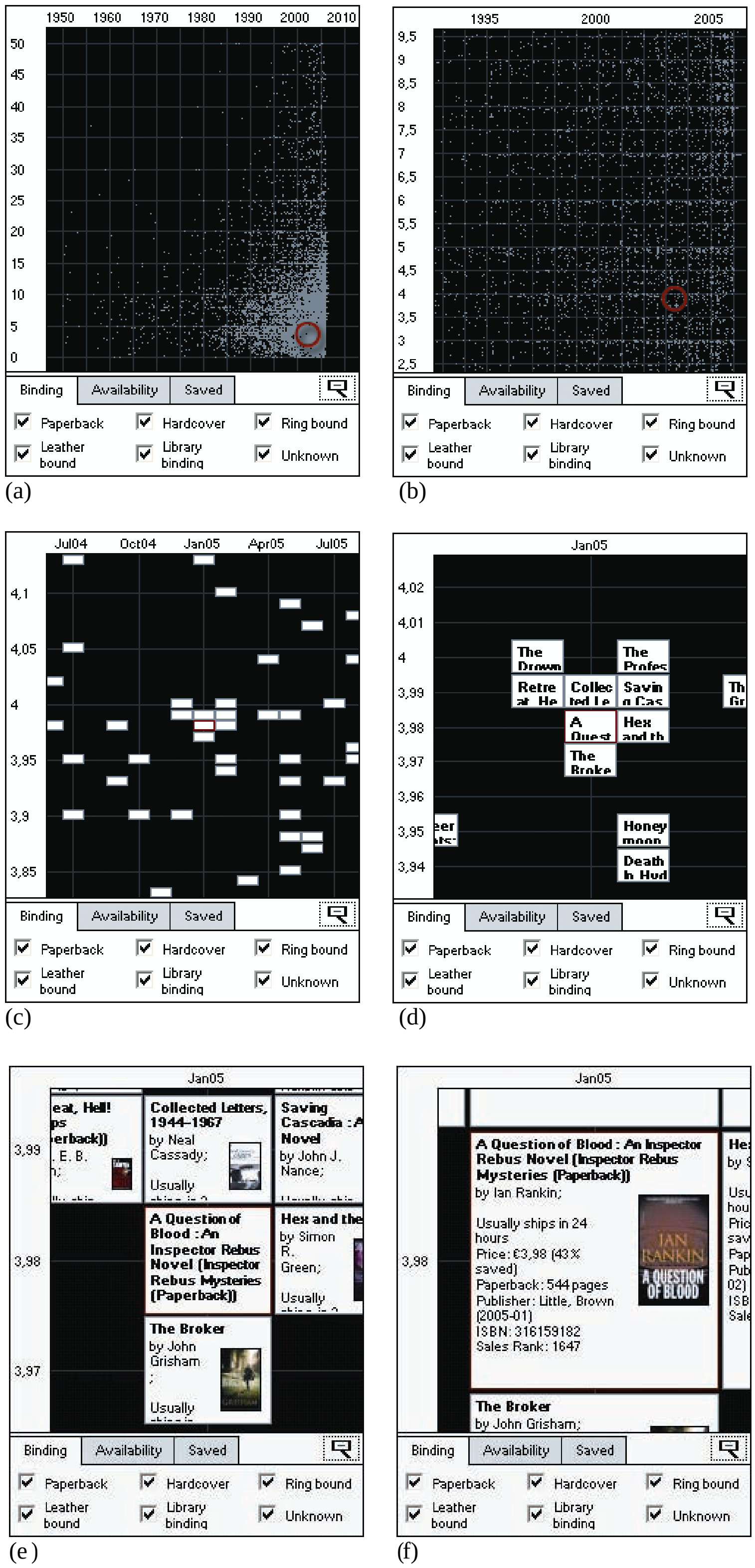
Interfaces

To examine the effect that a distortion strategy may have on the usability of a zoomable starfield display on PDAs, we implemented a fisheye view and a third design iteration of the detail-only prototype design. One requirement was that the applications should scale to larger information spaces of several thousands of items (in this case we used a book database). Moreover, dynamic query modules were added to the design to support data filtering. These tabbed controls at the bottom of the screen were used to provide the participants of the usability test with a more realistic search scenario, and to emulate different levels of information density. In both applications the X-axis represents the year of publication and the Y-axis the sale price.

Compared to the previous design, several interaction and visualization details of the detail-only interface have been improved to support scalability. At the default scale, the scatterplot contains thousands of 1-pixel representations (Figure 5a). Moreover, since the screen holds fewer pixels than there are attribute units of the scatterplot dimensions, there may be thousands of items that are completely overlapped by other items. To make each item easily and accurately accessible by the users, the zoom animation, which is still triggered after a pen-down threshold of 150 milliseconds, has been redesigned to follow a two-step algorithm: (1) only the scatterplot pane is magnified, that is, the size of all the representations relative to each other remains unchanged, but the information space is stretched like a rubber sheet. At the same time, the focus object, which has been selected based on pen-proximity, moves to the center of the screen. Together, this causes a decomposition effect in which clusters of in-focus representations scatter, while other representations drift out of sight (Figure 5b; the circles added to the screenshot denote the position of the focus item). (2) once all clusters have been resolved, meaning that each gray pixel represents only items with the same attribute values, the zoom animation also takes effect on the representations. Gray pixels grow to small rectangles (Figure 5c) and then to record cards that display information about the books they represent (Figures 5d-f). The transition between the two steps of the algorithm is fluent and automatic. No extra action, besides holding down the pen, is required from the users. One thing to note about the scatter effect is that it only aims to resolve clusters of items with different attribute values. To resolve overlappings of items with the same values for the scatterplot dimensions, a zooming approach for multiple data points as discussed in (Büring & Reiterer, 2005) would be a reasonable and consistent solution.

Zooming in from the default scale takes 1.8 seconds to reach the highest magnification level, and 1.3 seconds to zoom back out. The zoom is at its maximum when the record card in focus fills about 50% of the diagram size, as in Figure 5f. Users can easily switch the focus to other record cards in the view by simply tapping them. The newly selected focus item is centered in a quick animation. Panning is implemented as *sliding*, a

Figure 5. Exploring 7,500 books on a PDA



rate-based scrolling mechanism controlled by a pen gesture. Users tap the display and drag the pen in the direction they want to move the viewport (this is the opposite of dragging the information space). The further they drag the pen, the more the animated view movement accelerates, and vice versa. The benefit of sliding is that long distances can be covered very quickly. A similar panning approach has also been discussed and user-tested in (MacKay et al., 2005). To distinguish between zoom and pan attempts, we use a distance threshold of 5 pixels. A zoom operation is initiated if users hold the pen down for more than 150 milliseconds without moving it more than 5 pixels, otherwise the input is interpreted as a panning gesture.

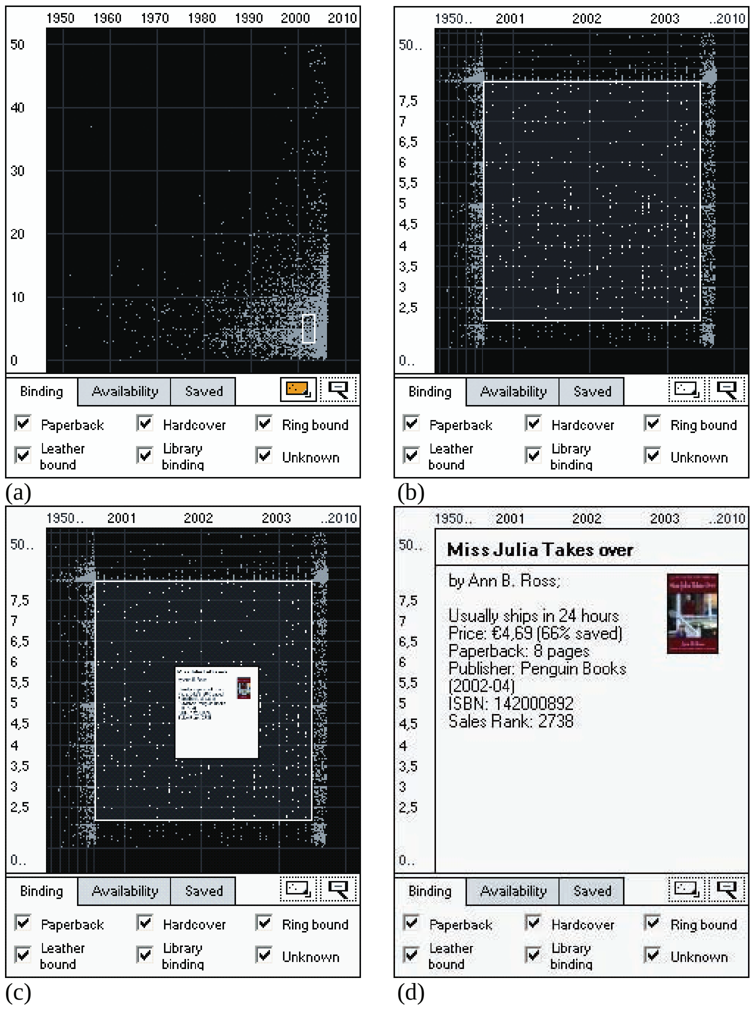
Another redesign aspect targets the record card presentation. To truly base the ratio of overview and detail information on the degree of zooming the users perform, the layout of the record cards changes as fluently and smoothly as the scale. First the application tries to display as much of the book title as possible, assuming that this is the most important portion of the content available (Figure 5d). In the case where vertical space is still left, it is filled with the scaled book cover. Any remaining horizontal space next to the cover is used to show additional textual details such as the author's name, shipping information, and so forth (Figure 5e). Much effort has also been spent on improving the scatterplot labeling. In the former versions, the labels remained static and, just like items that are not in focus, they moved out of the viewport during zooming. In the redesign, labels are rapidly and continuously updated to mirror the current level of view granularity. For instance, in the default scale the X-axis is subdivided into decades and the Y-axis into units of EUR 5 (Figure 5a). While zooming in, the labels and grid lines drift apart. The gap is used to display new grid lines that, as soon as there is sufficient display space available, are equipped with labels of the next-smaller unit measure (e.g., Figure 5c). When the scale has reached its maximum, each scatterplot dimension is displaying its smallest attribute unit. The predefined granularities are 10 years, 5 years, year, and quarter for the X-axis and 5 euro, 1 euro, 10 cent, 5 cent, and 1 cent for the Y-axis.

The fisheye interface makes use of a discrete interface distortion. To prune visual clutter, users first tap the box-icon button below the diagram and then draw a bounding-box denoting the focus region (Figure 6a). Upon lifting the pen, the system responds by smoothly centering and magnifying the boxed region in 600 milliseconds to about 75% of the diagram size (Figure 6b). This results in a layout in which more display space is available for the focus region (items drift apart, clutters can be resolved), while objects that are not in focus are still visible but are allocated less space. The preservation of context has the advantage that users can make informed decisions about where to navigate; they may, for instance, prevent themselves from panning into regions where there are no items. Users can also directly zoom into a context region by redrawing the bounding box to include a portion of the regions surrounding the current focus. To achieve this with an undistorted ZUI, users would need to previously zoom out before adjusting the focus.

Depending on the given information density, users may have to apply the fisheye zoom recursively to isolate items. An important property of the distortion applied is that it does not affect parallelism between lines. In contrast, a radial distortion may for instance have significantly increased the cognitive load to map items to the diagram labels. As with the detail-only application, the scatterplot labels are continuously updated during the magnification animation. However, to avoid confusing overlappings, the context regions are only labeled with the start or end unit of the axis. To return to a previous distortion setting, or to undistort the interface, users tap, or tap and hold the button with the magnifying glass and the minus sign. Due to the compressed information space in this interface, precise panning is more important than fast panning. Hence, panning is implemented as drag&drop; users move items in and out of the focus by dragging the diagram pane with the stylus.

While the interface distortion allows users to resolve visual clutter, it does not reveal any detail information about book items. To shift from overview to content data, users tap the item they

Figure 6. Fisheye interface: (a) selecting a focus region, (b) distortion result, (c) item zooms to full-size, (d) detail view



are interested in. Again, they are not required to accurately tap the 1-pixel representation; it is sufficient to tap in the item's vicinity. The selected object zooms to full screen in 250 milliseconds. During the animation, the gray pixel grows to the familiar record card representation, showing all detail information about the book along with its cover (Figures 6c, d). On tapping anywhere on the record card, the item zooms back to its original size and position in the scatterplot. This approach is different to a conventional pop-up mechanism in that it leads the user's eye. The animation gives a visual cue about where the item comes from, and where it goes back to. Hence, users can quickly access the content of different items, without the need to reorientate after each zoom-out operation. The detail zoom can be applied to all visible items, regardless of whether they are located in the focus or in a context region, or whether the interface is currently distorted or not.

Evaluation Abstract

We conducted an experiment in which 24 participants were asked to solve retrieval tasks using either of the two interfaces. Search tasks were based on an information space of 7,500 book items. While we tested our highly animated applications successfully with a few hundred items on HP and Dell PDAs, current mobile devices are not yet powerful enough to cope with thousands of data records. Thus, we simulated a standard PDA interface (240 x 320 pixel) on a stylus-operated Wacom Board connected to a 3 GHz Pentium 4 PC with 1 GB RAM. This apparatus allowed us to use pen-interaction while also being provided with sufficient processing power. To run on the test machine, both applications were ported from the compact framework to the full .Net framework. The experiment is fully reported in (Büring et al., 2006b).

The results of the usability tests were rather surprising. Due to the integrated overview, we had expected that the fisheye view would result in quicker navigation and less unnecessary exploration (see also Schaffer, Zuo, Greenberg, Bartram, Dill, Dubs, et al., 1996). However, it took users about the same time to complete 10 tasks with both interfaces (623.8 seconds for the detail-only interface compared to 612.4 seconds for the fisheye interface, $F(1,22) = 0,002$, p=n.s.). Considering that the fisheye interface required far fewer actions, we must conclude that these actions took more time to execute. Hence, it seems that drawing a bounding box and then accessing items one-by-one is cognitively more demanding than the more comprehensive zooming approach in the detail-only interface.

Regarding preference, 20 subjects preferred the fisheye interface and only three the detail-only interface ($X^2(1,N=23)=12.565, p<0.001$). This difference is highly significant. User statements suggest that the result is due to the better orientation features and the more precise navigation offered by the fisheye interface. This finding partly contradicts related research. For instance, two studies that compared overview+detail interfaces to a fisheye view found that although the fisheye improved user performance for many task types, subjects were still clearly in favor of the alternative interface (Baudisch et al., 2004; Hornbæk & Frøkjær, 2001). While the rather artificial distortion of the fisheye view may have discouraged users in those experiments, we hypothesize that in our case, the fisheye benefited from the abstract representation of the diagram. The distortion integrates very well into the scatterplot layout. This may be quite different for domains such as maps, in which a higher degree of fidelity to the standard layout is essential (Plaisant et al., 1995).

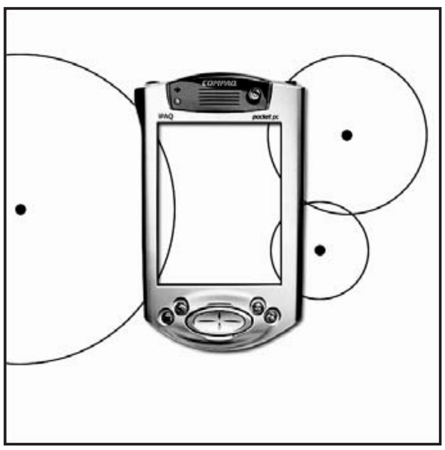
The preference for the fisheye view is also influenced by the users' preference for the more conventional drag&drop mechanism over the sliding technique (16:7, $X^2(1,N=23)=3.522$, $p=0.061$, n.s.). We think the reasons for this result are two-fold. First, all participants were used to drag&drop as it is supported by all major window systems. Sliding, on the other hand, was a highly unfamiliar technique. The other reason relates to the clipping approach in the detail-only prototype. Even though sliding enables users to pan at high speed in the information space, this feature does not compensate for the lack of overview information. If users do not know in which direction to move, speed is hardly an advantage. In addition, a drawback of the technique is that when users are sliding very quickly and pass an item, it appears only as a quick flicker of the screen. Thus, it takes users some time to react, and then they have to back-track to find the item. Comments suggest that our participants found this procedure rather difficult.

OUTLOOK

As already mentioned, the usability of the fisheye may depend quite strongly on the type of information space displayed. When applied to map-based starfields, users are likely to become irritated by distorted landmarks such as streets or buildings. The detail-only interface featuring the geometric semantic zoom preserves the standard layout and thus, may be better suited for geographic application domains. However, the interface suffers from a lack of overview information, and is also in need of navigation features that allow the user to move long distances quickly and precisely. To overcome these problems, we suggest that the interface should be enhanced with two techniques proposed by previous research.

Map-based starfields usually feature points-of-interest (POI), that is, locations that are emphasized by drawing a symbol on top of the map. POIs may, for instance, be gas stations or hotels in a navigation system. Given the detail-only interface, the overview could be significantly improved by visualizing off-screen objects. This can be accomplished by using a technique called Halos (Baudisch & Rosenholtz, 2003). Halos are rings that are drawn around clipped POIs, and are just large enough to reach into the current viewport (see Figure 7). Based on the halo curvature, users are able to estimate the distance and direction of off-screen items. When moving the viewport, the rings are rapidly updated. The authors evaluated Halos in comparison to conventional arrows. They found that for navigation tasks, Halos led to a significant timesaving of up to 33%. The downside is that Halos do not scale well. Even a small number of POIs can cause Halos to clutter and thus lessen the technique's usability. To improve clarity, relevance thresholds have been suggested to limit the number of rings to be displayed (Baudisch & Rosenholtz, 2003). However, this technique assumes a prior definition of relevance by the user. Hence, additional interaction widgets are required that must be designed such that the simplicity of the original visualization is not hampered.

Figure 7. Halos are rings that visualize off-screen points-of-interest



Once Halos are displayed, the system must provide a way to quickly navigate to the corresponding off-screen locations. Due to visual blur, rate-based scrolling techniques, such as sliding, have been found hard to control. Separate zoom and pan operations are also not ideal, since they are slow and tedious to perform. A solution for this problem may be speed-dependent automatic zooming (SDAZ) (Igarashi & Hinckley, 2000). SDAZ couples scroll speed with scaling and thus combines zooming and panning into a single operation. When scrolling quickly, the display zooms out, and when reducing scroll speed, the display zooms back in. In this way, visual flow remains quite constant, which makes it easier for the users to navigate at high speed at the same time as they are being provided with an overview of the information space. While a preliminary informal evaluation for document and map navigation showed mixed results, more recent user tests revealed a clear advantage of SDAZ when compared to conventional scroll, pan, and zoom methods (Cockburn & Savage, 2003; Cockburn, Savage, & Wallace, 2005a; Cockburn, Savage, & Wallace, 2005b) . Another study applied SDAZ to a small screen that mimicked the standard PDA resolution (Jones, Jones, Marsden, Patel, & Cockburn, 2005). The novel zoom interaction had the effect of worsening task completion times and from that, the authors concluded that the small screen space had reduced the impact of SDAZ. However, since the study lacked a control group, more research must be conducted to clarify this point. In addition, participants controlled the interface via a computer mouse, which is not a standard input mechanism for devices that typically feature a small screen. Thus, it would be interesting to evaluate whether pen and mouse interactions have different effects on the usability of SDAZ.

Another topic for further research may be to examine the scalability of zoomable starfield displays in terms of screen size. The interfaces developed not only meet the requirements of mobile devices, but could also provide a valuable solution for presenting very large sets of data on desktop computers and wall-sized displays.

CONCLUSION

Several interaction strategies to improve the usability of starfield displays on PDAs have been implemented and evaluated. While smooth geometric and semantic zooming provided an intuitive metaphor for exploring information spaces, users still showed difficulties in preserving their orientation while navigating. An interface featuring a separate overview window was not able to improve user satisfaction and moreover, due to the small size of the control and the cognitive costs of visual switching, it worsened task completion times. The problem of visual switching could be avoided by introducing a fisheye distortion algorithm. Context regions are not clipped but are preserved by being contracted. More space can then be allocated to the items in focus. While task completion times remained similar to the default smooth-zooming starfield, users significantly preferred the fisheye view. This is an important result that may encourage designers to employ distortion strategies when displaying abstract information spaces on small screens. However, for less abstract data such as maps, the detail-only interface featuring the geometric semantic zoom may still provide the better solution, especially when enhanced by techniques such as speed-dependent automatic zooming and Halos.

ACKNOWLEDGMENT

This work was supported by the DFG Research Training Group GK-1042 "Explorative Analysis and Visualization of Large Information Spaces."

REFERENCES

Ahlberg, C. (1996). Spotfire: An information exploration environment. *SIGMOD Rec.*, *25*(4), 25–29.

Ahlberg, C., & Shneiderman, B. (1994). Visual information seeking: Tight coupling of dynamic query filters with starfield displays. In *Chi '94 Proceedings of the sigchi conference on human factors in computing systems* (pp. 313–317). New York: ACM Press.

Baudisch, P., Good, N., Bellotti, V., & Schraedley, P. (2002). Keeping things in context: A comparative evaluation of focus plus context screens, overviews, and zooming. In *Chi '02 Proceedings of the sigchi conference on human factors in computing systems* (pp. 259–266). New York, NY: ACM Press.

Baudisch, P., Lee, B., & Hanna, L. (2004). Fishnet, a fisheye web browser with search term popouts: A comparative evaluation with overview and linear view. In *Avi '04 Proceedings of the working conference on advanced visual interfaces* (pp. 133–140). New York, NY: ACM Press.

Baudisch, P., & Rosenholtz, R. (2003). Halo: A technique for visualizing off-screen objects. In *Chi '03 Proceedings of the sigchi conference on human factors in computing systems* (pp. 481–488). New York, NY: ACM Press.

Bederson, B., & Meyer, J. (1998). Implementing a zooming user interface: Experience building pad++. *Softw. Pract. Exper.*, *28*(10), 1101–1135.

Bederson, B. B., Clamage, A., Czerwinski, M. P., & Robertson, G. G. (2004). Datelens: A fisheye calendar interface for pdas. *ACM Transactions on Computer-Human Interaction*, *11*(1), 90–119.

Björk, S., Holmquist, L. E., Redström, J., Bretan, I., Danielsson, R., Karlgren, J., et al. (1999). West: A web browser for small terminals. In *Uist '99 Proceedings of the 12th annual acm symposium on user interface software and technology* (pp. 187–196). New York, NY: ACM Press.

Burigat, S., & Chittaro, L. (2005). Visualizing the results of interactive queries for geographic data on mobile devices. In *Gis '05 Proceedings of the acm international workshop on geographic information systems* (pp. 277–284). New York, NY: ACM Press.

Büring, T., Gerken, J., & Reiterer, H. (2006a). Usability of overview-supported zooming on small screens with regard to individual differences in spatial ability. In *Avi '06 Proceedings of the working conference on advanced visual interfaces* (pp. 233–240). New York, NY: ACM Press.

Büring, T., Gerken, J., & Reiterer, H. (2006b, September). User interaction with scatterplots on small screens: A comparative evaluation of geometric-semantic zoom and fisheye distortion. *IEEE Transactions on Visualization and Computer Graphics (Proceedings Visualization / Information Visualization 2006), 12*(5), 829–836.

Büring, T., & Reiterer, H. (2005). Zuiscat: Querying and visualizing information spaces on personal digital assistants. In *Mobilehci '05 Proceedings of the 7th international conference on human computer interaction with mobile devices & services* (pp. 129–136). New York, NY: ACM Press.

Card, S. K., Mackinlay, J. D., & Shneiderman, B. (Eds.). (1999). *Readings in information visualization: Using vision to think*. San Francisco: Morgan Kaufmann Publishers.

Cockburn, A., & Savage, J. (2003). Comparing speed-dependent automatic zooming with traditional scroll, pan, and zoom methods. In *People and computers xvii: British computer society conference on human computer interaction* (pp. 87–102).

Cockburn, A., Savage, J., & Wallace, A. (2005a). Comparing automatic and manual zooming methods for acquiring off-screen targets. In *People and computers xix: British computer society conference on human computer interaction* (pp. 439– 454).

Cockburn, A., Savage, J., & Wallace, A. (2005b). Tuning and testing scrolling interfaces that automatically zoom. In *Chi '05 Proceedings of the sigchi conference on human factors in computing systems* (pp. 71–80). New York, NY: ACM Press.

Dix, A., & Ellis, G. (1998). Starting simple: Adding value to static visualisation through simple interaction. In *Avi '98 Proceedings of the working conference on advanced visual interfaces* (pp. 124–134). New York, NY: ACM Press.

Dunlop, M. D., & Davidson, N. (2000). Visual information seeking on palmtop devices. In *Proceedings of hci2000, volume 2* (p. 19–20).

Dunlop, M. D., Morrison, A., McCallum, S., Ptaskinski, P., Risbey, C., & Stewart, F. (2004). Focussed palmtop information access combining starfield displays and profile based recommendations. In F. Crestani, M. Jones, & S. Mizzaro (Eds.), *Proceedings of workshop on mobile and ubiquitous information access, lncs v2954* (p. 79–89). Springer.

Furnas, G. W. (1986). Generalized fisheye views. In *Chi '86 Proceedings of the sigchi conference on human factors in computing systems* (pp. 16–23). New York, NY: ACM Press.

Hornbaek, K., Bederson, B. B., & Plaisant, C. (2002). Navigation patterns and usability of zoomable user interfaces with and without an overview. *ACM Trans. Comput. Hum. Interact., 9*(4), 362–389.

Hornbaek, K., & Frøkjaer, E. (2001). Reading of electronic documents: The usability of linear, fisheye, and overview+detail interfaces. In *Chi '01: Proceedings of the sigchi conference on human factors in computing systems* (pp. 293–300). New York, NY: ACM Press.

Igarashi, T., & Hinckley, K. (2000). Speed-dependent automatic zooming for browsing large documents. In *Uist '00: Proceedings of the 13th annual acm symposium on user interface software and technology* (pp. 139–148). New York, NY: ACM Press.

Jones, S., Jones, M., Marsden, G., Patel, D., & Cockburn, A. (2005). An evaluation of integrated zooming and scrolling on small screens. *Int. J. Hum. Comput. Stud.*, *63*(3), 271–303.

Karlson, A. K., Bederson, B. B., & SanGiovanni, J. (2005). Applens and launchtile: Two designs for one-handed thumb use on small devices. In *Chi '05: Proceedings of the sigchi conference on human factors in computing systems* (pp. 201–210). New York, NY: ACM Press.

Lamping, J., Rao, R., & Pirolli, P. (1995). A focus+context technique based on hyperbolic geometry for visualizing large hierarchies. In *Chi '95 Proceedings of the sigchi conference on human factors in computing systems* (pp. 401–408). New York, NY: ACM Press/Addison-Wesley Publishing Co.

MacKay, B., Dearman, D., Inkpen, K., & Watters, C. (2005). Walk 'n scroll: A comparison of software-based navigation techniques for different levels of mobility. In *Mobilehci '05 Proceedings of the 7th international conference on human computer interaction with mobile devices & services* (pp. 183–190). New York, NY: ACM Press.

Mackinlay, J. D., Robertson, G. G., & Card, S. K. (1991). The perspective wall: Detail and context smoothly integrated. In *Chi '91 Proceedings of the sigchi conference on human factors in computing systems* (pp. 173–176). New York, NY: ACM Press.

Perlin, K., & Fox, D. (1993). Pad: An alternative approach to the computer interface. In *Siggraph '93 Proceedings of the 20th annual conference on computer graphics and interactive techniques* (pp. 57–64). New York: ACM Press.

Plaisant, C., Carr, D., & Shneiderman, B. (1995). Image-browser taxonomy and guidelines for designers. *IEEE Software*, *12*(2), 21–32.

Schaffer, D., Zuo, Z., Greenberg, S., Bartram, L., Dill, J., Dubs, S., et al. (1996). Navigating hierarchically clustered networks through fisheye and full-zoom methods. *ACM Trans. Comput. Hum. Interact.*, *3*(2), 162–188.

Shneiderman, B. (1996). The eyes have it: A task by data type taxonomy for information visualizations. In *Vl '96 Proceedings of the 1996 ieee symposium on visual languages* (pp. 336–343). IEEE Computer Society.

Shneiderman, B., & Plaisant, C. (2005). *Designing the user interface* (4th ed.). Pearson/Addison-Wesley.

Spence, R. (2001). *Information visualization*. Addison-Wesley.

Spence, R., & Apperley, M. (1982). Data base navigation: An office environment for the professional. *Behaviour and Information Technology*, *1*(1), 43–54.

Tufte, E. R. (1983). *The visual display of quantitative information*. Cheshire, CT: Graphics Press.

Williamson, C., & Shneiderman, B. (1992). The dynamic homefinder: Evaluating dynamic queries in a real-estate information exploration system. In *Sigir '92 Proceedings of the 15th annual international acm sigir conference on research and development in information retrieval* (pp. 338–346). New York: ACM Press.

KEY TERMS

Dynamic Queries: Graphical widgets, such as sliders and checkboxes, used to formulate queries in a direct manipulation interface.

Focus+Context Interface: An interface that uses distortion to integrate both focus and context in a single view.

Halos: A visualization technique to indicate the position and distance of off-screen objects. Around each clipped object a ring is drawn, which is just large enough to reach into the current viewport.

Overview+Detail Interface: A multiwindow interface, where usually one window is used to present details while the other one gives an overview of the entire information space.

Speed-Dependent Automatic Zooming (SDAZ): SDAZ couples rate-based scrolling with scaling and thus combines zooming and panning into a single operation.

Starfield Display: A complex search interface that visualizes abstract data as an interactive scatterplot diagram enhanced with zooming and filter capabilities.

Zoomable User Interface (ZUI): A spatial 2.5-D interface, in which data objects are organized in space and scale. Users navigate this space by performing zooming (changing the scale) and panning (movement at constant scale) operations.

Chapter XXXV
Projected Displays of Mobile Devices for Collaboration

Masanori Sugimoto
University of Tokyo, Japan

ABSTRACT

Mobile devices have so far been personal tools. With their evolution of increased functionality, however, these devices have begun to be used in a shared fashion by multiple people. This chapter discusses techniques allowing multiple people to share mobile devices by projecting their displays and conducting intuitive manipulations on them. The chapter first shows overviews of systems and technologies related to location-aware projection and several interaction techniques. Then, a system called Hotaru that implements intuitive manipulation techniques on projected displays of multiple mobile devices is described. Hotaru allows a user to annotate or rotate a picture or a document on a projected display by using his finger and intuitively to transfer a file between multiple devices by overlapping their projected displays. User studies of Hotaru indicated that the proposed manipulation techniques could support multiple people in a single location in conducting their tasks. Research issues on projected displays of mobile devices are raised.

INTRODUCTION

Mobile devices (PDAs, cellular phones, etc.) have rapidly penetrated into our society, and many people use them in their daily lives. For example, in Japan, the number of subscribers of cellular phones is 93 million (as of the end of July 2006), which is more than three-quarters of the total Japanese population (Telecommunication Carrier Association, 2006).

One of the recent trends of mobile devices is multifunctionality: they are used not only as a communication tool with another person or as a personal scheduler, but also as a Web browser, a digital video camera, a game console, a music player, a television, a GPS device, electronic wallet, and so on. This trend makes the difference between a mobile device and a personal computer smaller: a cellular phone or a PDA is taking the role of another "personal computer" retaining the feature of mobility, although their computational capability is still lower than desktop or notebook computers. Moreover, with the evolution and improvement of mobile devices in their functionality and performance, they have begun to be used by multiple people in face-to-face or colocated situations.

The following example shows how a mobile device is used in a colocated situation. A user takes a photo by using a digital camera built into his cellular phone, and wants to show the photo to all the people around him. However, because of the problem of screen real estate on a cellular phone, it is not easy for multiple people to look simultaneously at the photo on his cellular phone. When he tries to pass the photo to people who have requested it, he needs to conduct unintuitive and bothersome operations on his phone using a limited number of keys; for example, by sending the photo to the other people using e-mail, or transferring the photo using infrared communication between two phones brought close together.

One method for solving this problem is projecting a display of a mobile device in order to make it sharable by multiple people and to allow them to conduct intuitive manipulations on the display. For example, if a projector mountable onto a mobile device is available, a user can easily make a projected display of the device appear in any size on any convenient surface such as a wall, a floor, or a table. When multiple users surrounding the display can conduct manipulations it, such as annotation, file selection, file transfer, and so on in an intuitive way, they can collaborate with each other more smoothly and easily.

Because of weight and power-consumption problems, a projector mountable onto a mobile device and usable for sharing its display among multiple people is not currently available. The Canesta Keyboard (Roeber, Bacus, & Tomasi, 2003) is a monochrome short-distance projection system designed to be attached to a PDA, and used to project a personal virtual keyboard (therefore, its projected display is too small to be shared by multiple people). According to recent news, however, research into portable projectors is in progress, and units mountable onto mobile devices will become available in the near future (New York Times, 2004). On the other hand, a video camera mountable onto a mobile device is commercially available now (e.g., mobile phones with built-in video cameras) and can be used for recognizing users' manipulations on its projected display.

In this chapter, therefore, we propose a system called *Hotaru* ("Firefly") that allows users to conduct intuitive manipulations on projected displays of mobile devices (PDAs in this study) by utilizing currently available technologies. In *Hotaru*, the displays of PDAs are projected through a projector attached to a ceiling based on their 3-D positions and orientations. A stereo camera installed on the ceiling is used to capture and track the positions and orientations of the PDAs. When a user holding his PDA moves it in 3-D space, he can change the location, size, and shape of its projected display accordingly.

Users' manipulations on projected displays using their fingers are recognized through a video camera attached to a mobile device. *Hotaru* allows users to conduct mouse-style operations, such as click, double-click, drag, and to annotate or rotate documents or images (Miyahara, Inoue, Tsunesada, & Sugimoto, 2005). *Hotaru* also implements a novel and intuitive file-transfer technique between PDAs by overlapping their projected displays (Sugimoto, Miyahara, Inoue, & Tsunesada, 2005).

The chapter is organized as follows: The next section discusses related work to *Hotaru*. Then, the system configuration of *Hotaru* and its technological details are described. In the Intuitive Manipulation Techniques section, several manipulation techniques for enhancing colocated collaboration and their user interface issues are shown. In the User Studies section, evaluations of *Hotaru* are discussed. Finally, research issues to be investigated for a mobile device with a projector and a camera are raised.

RELATED WORK

Hotaru is related to several research topics such as mobile and ubiquitous computing, gesture recognition based on computer vision techniques, and so on. In this section, studies on location-aware projection techniques/systems and intuitive interaction techniques on projected displays, which are most related to *Hotaru* in a technical sense, are introduced. iLamps (Raskar, van Baar, Beardsley, Willwacher, Rao, & Forlines, 2003) provides users with adaptive projection techniques using a cluster of handheld projectors. iLamps identifies the locations and orientations of multiple projectors, and creates a seamless and consistent image over a projected surface (e.g., planar, spherical, etc.) by correcting the overlapping image given by the projectors. In Pingali et al. (Pingali, Pinhanez, Levas, Kjeldsen, Podlaseck, Chen, & Sukaviriya, 2003), the concept of a "steerable interface" and its implementation are described. The proposed system (ED: Everywhere Display) uses an LCD projector fixed to a ceiling and can project a display on any surface (e.g., a wall or a floor in a room) by tracking a user's head position and controlling the angle of a mirror attached to the projector. In BurningWell (Summet & Sukthankar, 2005) and the systems presented in Lee et al. (Lee, Dietz, Maynes-Aminzade, Raskar, & Hudson, 2005), light sensors placed on a target surface capture a series of gray-coded binary patterns and identify the corresponding pixel coordinates in the projected image. The systems then calibrate the projected images to fit the target surface or change their locations by tracking a moving sensor.

PaperWindows (Holman, Vertegaal, Altosaar, Troje, & Johns, 2005) tracks the 3-D position and shape of a real paper through its attached makers and multiple cameras. It then generates a warped image to be projected correctly onto the paper from a projector placed at a fixed position. PaperWindows allows users to manipulate digital information intuitively through their hand and finger gestures utilizing physical features of paper, such as holding, flipping, rubbing, and so on. HyperPalette (Ayatsuka, Matsushita, & Rekimoto, 2000) allows a user to conduct intuitive operations with his PDA. A user can bring a photo projected onto a table into his PDA (scoop), or drop a photo from the PDA onto the table (spread) by tilting and moving the PDA. Sukthankar proposes a projector–camera system for automatically calibrating projected images (Sukthankar, Stockton, & Mullin, 2001). The system is used for slide presentation tasks by tracking a user's pointing device (e.g., laser pointer). Augmented Surfaces (Rekimoto & Saito, 1999) allows users to drag files, such as documents or images, to be shown on a desktop computer, a table, or a wall, by using a laser pointer. Users can easily share these files with other users and bring them into their own personal computers.

PROJECTING AND RECOGNIZING DISPLAYS OF MOBILE DEVICES

Design Requirement

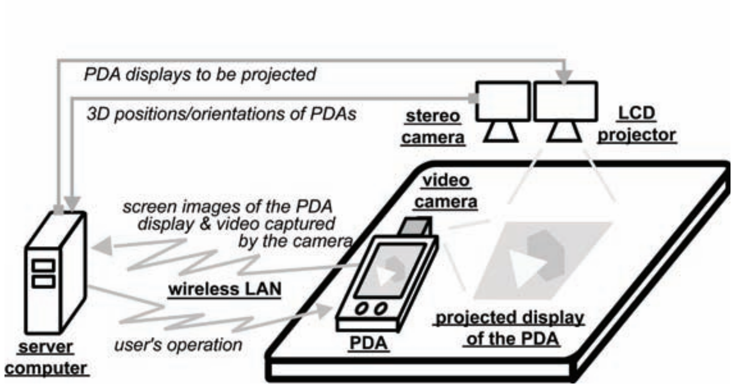
In this section, the generation of a projected mobile device display and recognition of the display are described. In order to project a display without a projector attached to the mobile device, the following requirements are examined.

1. 3-D positions and orientations of mobile devices are automatically identified.
2. Their displays are projected based on the positions and orientations using existing projectors.

In order to fully satisfy requirement 2, a special apparatus to rapidly and accurately control the positions and orientations of LCD projectors will be necessary. However, it is almost impossible to implement such an apparatus, for example, a moving platform for mounting multiple pan-tilt projectors that can instantaneously change their positions and orientations by following the 3-D movement of multiple mobile devices.

On the other hand, it is possible to investigate the idea of projected displays of mobile devices and intuitive manipulation techniques on them by limiting the surface areas to be projected onto,

Figure 1. System configuration of Hotaru



without the use of special apparatus. Therefore, in this study, we decided to develop a prototype version of *Hotaru* without controlling the position and orientation of projectors. In *Hotaru*, the projected displays of mobile devices are shown anywhere on a specified table through a projector installed on the ceiling.

In this version of *Hotaru*, a PDA is used as a mobile device, because the software development environment for PDAs is much better than that for cellular phones. Because of the limited computational capability of current models of PDAs, an image processing module for recognizing users' manipulations on projected displays is executed on a server computer as shown in Figure 1.

Recognition of 3-D Position and Orientation

As for requirement 1 in Section 3.1, a marker with infrared (IR) LEDs and a stereo camera are used for recognizing the 3-D position and orientation of a user's PDA. The IR LEDs on a circuit board are arranged to form an isosceles triangle (*A*, *B*, and *C* in Figure 2). The position of the PDA is represented as $P = (x_p, y_p, z_p)$, which are the coordinates of the centroid of the triangle. The orientation of the PDA is calculated by the sum of the vector $\overrightarrow{BA}$ and $\overrightarrow{CA}$, and represented as $\mathbf{v} = (x_o, y_o, z_o)$. Different blinking patterns are assigned to LEDs at the vertex A for identifying the individual PDAs.

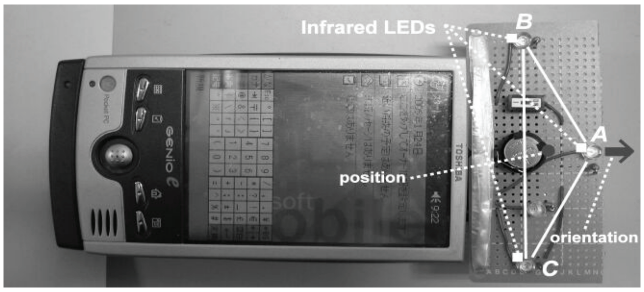
Experimental results of evaluating this method have proved that the position and orientation recognition errors are less than six centimeters and five degrees, respectively, when the stereo camera (ViewPLUS Bumblebee) is installed on the ceiling 2.5 m above the floor (The errors were evaluated by the distance between the estimated and real positions, and by the angle formed by the estimated and real orientation vectors, respectively). In the current implementation, it takes less than a second to correctly recognize individual blinking patterns of LEDs. This means that a user is required to hold his PDA steady for 1 second so that *Hotaru* can identify the PDA successfully.

Projecting Displays of Mobile Devices

Figure 3 shows how the projected display of a PDA is determined. In *Hotaru*, a projector is assumed to be mounted at the point *P* (the position of a PDA as discussed in Section 3.2) and pointed in the direction of the vector **d** (perpendicular to the PDA screen). Let the projection plane be π, and the intersection point of the line passing through *P* in the direction of vector **d** and the plane π be *C*, and the quadrangular pyramid generated by the light rays of the virtual projector at the point *C* be **Q**. In this case, the projected display of the PDA is given by the intersection of the plane π with the quadrangular pyramid **Q**.

In order to update the projected display of the PDA held and moved in 3-D space by a user, the

Figure 2. Infrared LEDs for position and orientation identification



rapid recognition of its 3-D position and orientation, which requires an increase in the frame rate of the stereo camera, is necessary. In the current implementation of *Hotaru*, the frame rate is set to 15 fps. Because of the errors, as described in Section 3.2, a projected display of a PDA fluctuates and is not stable when *Hotaru* uses its estimated position and orientation data directly. Therefore, *Hotaru* calculates the PDA's current positions and orientations by averaging the recent 10 consecutive location and orientation measurements. This causes a tracking delay in the projected display of a PDA when it is moved quickly. However, by making the projected display more steady, users feel less frustrated and the number of manipulation errors is reduced.

Recognition of Projected Displays

Hotaru uses a video camera (SONY AVC666SN, 640 × 480 pixels) mounted onto a PDA and recognizes its projected displays as follows.

1. Extract contours and vertices of projected displays from a captured image through the camera
2. Identify projected displays of individual PDAs

Figure 4 shows the process of extracting contours and vertices of a projected display. In order to detect multiple projected displays, a wide-angle lens is mounted on the camera. *Hotaru* first performs a correction of the image distortion caused by the wide-angle lens, and applies the Canny edge detector (Trucco & Verri, 1998) for detecting contours and vertices in the image. When *Hotaru* detects four vertices of individual projected displays, it can successfully recognize the displays. If *Hotaru* cannot find all the vertices because of occlusion by human hands or overlap of other projected displays (Figure 4(a)), it applies the Hough transform (Figure 4(b)) and estimates the unrecognized vertices in order to determine the regions of individual projected displays (Figure 4(c)). When *Hotaru* cannot identify the vertices using these methods, it shows users a message about the failure of the identification through another monitor display.

By using the four vertices of each projected display in the PDA's camera coordinates and those of the PDA screen in world coordinates (given through its 3-D position and orientation), a transformation matrix between the two coordinates is calculated (Trucco & Verri, 1998) after identifying the projected displays of individual PDAs, as discussed in the next section. This transformation matrix is used to map a user's finger position on each projected display to its PDA screen, as shown in a later section.

Identifying Projected Displays of Mobile Devices

Although projected displays are detected from an image captured by the PDA's camera, it is not sufficient because the correspondence between the projected displays and the individual PDAs has not

Figure 3. How Hotaru generates a projected PDA display

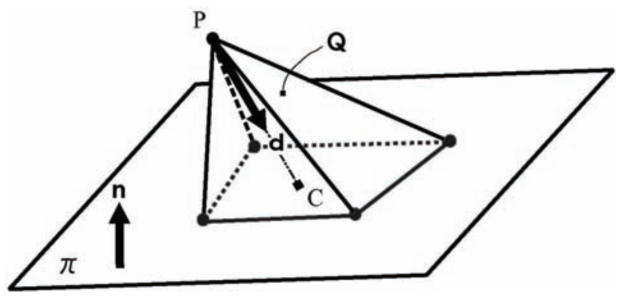
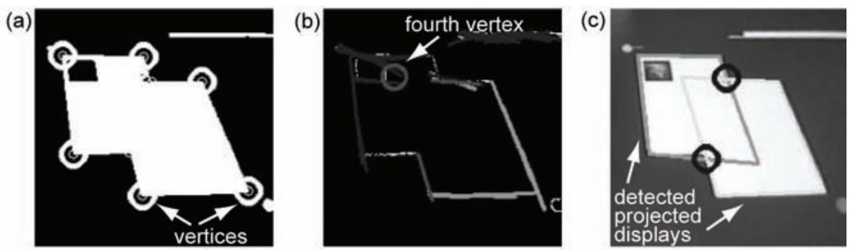


Figure 4. Extracting a projected display. (a) Three vertices of individual projected displays are determined. (b) The fourth vertex of each projected display is estimated. (c) The region of each projected display is recognized.



been determined. By estimating the locations of the projected displays based on the 3-D locations and orientations of individual PDAs, it may be possible to find the correspondence. However, if a portable projector mounted on the mobile device becomes available, such location and orientation information will be unnecessary. This means that this information should not be used for any purpose other than generating projected displays of PDAs. Therefore, finding the correspondence between PDAs and their projected displays is conducted by using the image of multiple projected displays captured by a camera mounted on each PDA.

We first tested similarity matching between an elicited projected display and the screen image of each PDA. However, this matching was not always successful, because (1) what was shown on a display were often only folders and icons, which were almost the same in all the projected displays, and (2) it was difficult to identify features of individual projected displays when they overlapped.

As one of the main purposes of the study was to examine the concept of intuitive manipulation methods on projected displays, a method that achieves more stable identification of multiple PDAs was used. In the current implementation of *Hotaru*, individual PDAs are given different-colored markers for their identification. In order to use a screen of a PDA as large as possible, the marker is projected outside of the projected display, as shown in Figure 5 or Figure 6. A problem with this method is that the number of PDAs that can be successfully identified is small (less than 10). To solve this problem, we are now investigating different methods for identifying multiple PDAs.

INTUITIVE MANIPULATION TECHNIQUES

Manipulation Techniques by Fingers

We first tried to recognize the manipulations by bare fingers as described in Wellner (1993). However, the recognition of bare fingers was extremely poor without using a special camera such as a thermo-infrared camera (Koike, Sato, & Kobayashi, 2001), which is not typically available for a mobile device. Therefore, in this version of *Hotaru*, we decided augment a user's finger with a red LED to increase the recognition performance (as shown in Figure 5).

Hotaru establishes the finger-pointing area as those pixels whose values are larger than a specified threshold of brightness. The position of the pointing area on a PDA screen is calculated by using the transformation matrix described in Section 3.4. *Hotaru* uses the Kalman filter (Welch, 2001) to estimate the next pointing area by using the current and previous pointing areas, when users conduct translation or drag operations, as explained next.

The following is a list of the operations. To identify whether a user has really touched a projected display and conducted a click operation with his finger is difficult. Therefore, two-dimensional moves of a user's finger and its dwell time are used for recognition of these operations.

- **Click:** When a user's pointing area has not changed for more than one second, and no folder, file, and icon has been selected, *Hotaru* recognizes that the user has conducted a click.
- **Double click:** When a user's pointing area has not changed for more than two seconds, and no folder, file, and icon has been selected, *Hotaru* recognizes that the user has conducted a double click.
- **Drag:** When *Hotaru* identifies a move of a user's finger after a click, it recognizes the move as a drag.
- **Release:** When *Hotaru* identifies that a user's finger has stopped for more than one second after a drag, it recognizes the stop as a release.
- **Cancel:** A user can cancel his current operation by quickly moving his finger away from the projected display.

The dwell time for a click, a double click, and a release was determined through several informal user studies. The participants in the studies commented that auditory feedback should be provided to let users know types of operations recognized by the system. Therefore, *Hotaru* uses a mouse-click sound as auditory feedback. For example, when *Hotaru* identifies that a user's finger has stopped for a second and recognizes the operation as a click, it emits the sound once. If *Hotaru* identifies that a user's finger has stopped one more second, and recognizes the operation as a double click, it emits the sound again. Users can know if their operations have been correctly identified through the auditory feedback.

Examples of Intuitive Manipulation on Projected Displays

Annotation

When multiple users share a document or a picture on a projected display, they can write comments or draw figures on it directly with their fingers. In Figure 5, a group of people share a map on a projected display, and one of them draws lines or marks on it with his finger. Such an annotation is useful in a colocated situation: when one person accesses a Web page to show a map by using his mobile device, the other people can easily suggest, recognize, or confirm the direction to their destination.

Figure 5. Annotation to a projected map

Figure 6. Rotation of a projected display

Rotation

When an image file in a PDA is projected on a table where multiple people sit around it, it is desirable to allow them to rotate the file so that each of them can easily view and understand the content of the image. As shown in Figure 6, a user can rotate an image about its center point through any angle, by clicking and dragging at the lower right corner of the image.

File Transfer by Overlapping Projected Displays

Figure 7 shows how users can transfer a file between multiple PDAs. In Figure 7(a), displays of two users' PDAs are projected onto a table. User A moves his PDA so that its projected display overlaps with that of user B's PDA (Figure 7(b)) and then drags an image file to be transferred to the overlapping region of their projected displays (Figure 7(c)). The overlapping region is also visualized on the screens of their PDAs. When user A releases the file, a popup window appears on the screen of user B's PDA to confirm whether the file_transfer is permitted.

If user B presses the "OK" button on the popup window, the file is transferred to his PDA as shown in Figure 7(d). It is also possible for users to write/draw comments on the overlapping region for "carbon-copying" as shown in Figure 7(c).

USER STUDIES

Overview

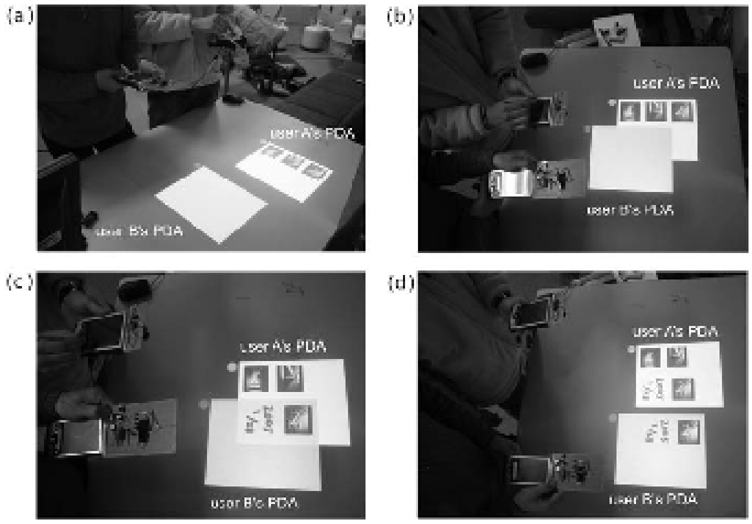
To evaluate the *Hotaru* system, the following two experiments were conducted.

- **Experiment 1:** Evaluations of how precisely *Hotaru* can identify a user's finger on its projected display.
- **Experiment 2:** Evaluations of the functions of *Hotaru*.

A user of *Hotaru* holds his PDA, and he/she or the other users conduct manipulations on its projected displays. Therefore, the precision ratio for identifying a finger position on the projected display may become worse than the ratio for identifying it in the case where the PDA is fixed. The purpose of Experiment 1 is to evaluate the differences of the precision ratios between these two cases.

In Experiment 2, the subjects were asked to conduct annotation, rotation, and file_transfer tasks as shown in Section 4.2. The purpose of this

Figure 7. File transfer between multiple devices by overlapping. (a) Displays of user A's PDA and user B's PDA are projected. (b) User B moves his PDA so that its projected display overlaps with that of user A's PDA. (c) User A drags a file to the overlapping region of their projected displays. (d) The file in user A's PDA is transferred to user B's PDA.



experiment is to clarify issues to be solved or improved through video analysis, questionnaires, and postexperimental interviews of the subjects.

Experiment 1

Four graduate students from our university participated in the experiment. Each of them was asked to touch a point randomly selected on the projected display with their finger. The experiment was conducted in the following two settings.

- **Fixed:** A PDA was fixed at a position 35 cm above a table to be used for projection
- **Handheld:** Another subject held a PDA at the same position as the fixed condition

Each subject was asked to touch a randomly selected point on the projected display 10 times under these two conditions (As one of the subjects performed the task only 9 times, the total number of touch trials by the subjects amounted to 39 in both settings).

Figure 8 shows a distribution of the distance between a randomly selected position and a subject's finger position recognized by *Hotaru*. The average error and standard deviation in the fixed setting were 6.6 pixels and 5.1 pixels, and those in the handheld setting were 8.3 pixels and 4.4 pixels, respectively. The recognition errors in both settings were mostly because of the estimation errors of the transformation matrix discussed in Section 3.4. In the handheld setting, another source of error was thought to be an unintentional movement or "jiggle" by the subject holding the PDA, which changed the camera and projected display locations.

However, as shown in Figure 8, the differences between these two settings were small in that the maximum error was less than 20 pixels. This means that even in the handheld setting, *Hotaru* can successfully identify user manipulations on projected displays when a size of an icon or a button to be manipulated by a user is large enough (for example, 40 × 40 pixels).

The reason for the small differences between the two settings seemed to be that (1) a subject tried to hold a PDA as steadily as possible so that he did not hinder tasks by another subject, and (2) as the movement of the camera and the associated projected display were synchronized, the location of the projected display in the image captured by the camera was stable (always located at the center of the image).

Experiment 2

Twelve subjects formed three pairs of two and two groups of three, and were asked to conduct the following tasks: annotation and rotation of an image file, and file transfer between multiple PDAs. The tasks lasted about 30 minutes for each pair and group.

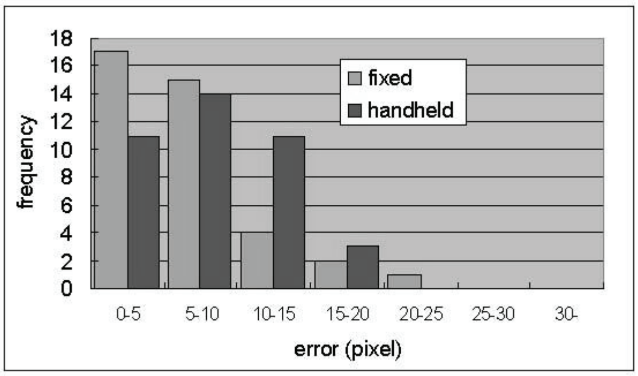
Positive (1–4) and negative (5–7) comments received from the subjects are summarized as follows.

1. Projected displays of PDAs were favored by the subjects. They could easily change the positions, shapes, and sizes of the displays as they liked, by moving their own PDAs.
2. *Hotaru* could effectively support multiple people in viewing pictures or documents, because it did not force them to look at a screen of another user's PDA over their shoulders.
3. Annotating and rotating files by finger was intuitive and useful.
4. Conducting file transfer tasks by overlapping projected displays of PDAs was much more intuitive and easier than other file transfer methods for mobile devices.
5. Slow responses to manipulations by finger were often irritating.
6. Recognizing user manipulations failed when a PDA was not held steadily enough.
7. It was desirable to identify who conducted which manipulation, in order for *Hotaru* to support collaborative tasks fully.

Comments 1–4 indicated that the idea of *Hotaru* was accepted by the subjects. The cause of Comment 5 was that the subjects had to stop the movement of their fingers and wait, in order to make a click, release, and double click operations recognizable by *Hotaru*. Similar reports are found in Zhang et al. (Zhang, Wu, Shan, & Shafer, 2001). As for comment 6, when a subject holding his PDA by one hand tried to touch his projected display with a finger of the other hand, the recognition often failed because he suddenly and unintentionally moved his PDA. Using inertial sensors (e.g., an accelerometer or a gyroscope) or optical flow analyses may be effective to reduce the influence of users' unintentional small movements of the PDA, and to fix the projected display at a specified location. As for comment 7, there are several possible solutions; for example, using visual or optical tags attached to fingers, in order for *Hotaru* to identify who has conducted which manipulation.

In the current implementation, *Hotaru* allows only a sender to conduct file-transfer tasks; that is, a sender first selects a file to be transferred,

Figure 8. Error distribution of pointing manipulations



and then releases it in a region overlapping with a projected display of another PDA. All the subjects requested that *Hotaru* should also allow a receiver to conduct file-transfer tasks by overlapping projected displays and selecting a file to be fetched. We will plan to improve and extend the functionality of *Hotaru* by examining the issues raised through the user studies.

RESEARCH ISSUES TO BE INVESTIGATED

In this chapter, new possibilities and applications for mobile devices have been explored: projection of mobile device displays and intuitive manipulation techniques on them. Although a small projector for a mobile device is not available at the moment, it is believed to be available in near future through the development of hardware technologies. In this section, the following research issues to be investigated using a mobile device with an attached projector and a video camera are raised.

- **Stable projection of a mobile device display:** Because of the jitter of a user's hand that holds a mobile device, its projected display fluctuates on a projected surface. It is desirable that the projected display of a mobile device be as stable as possible so that other users can easily recognize the contents shown on the display and conduct manipulations on it.
- **Calibration of a projected display:** One of the merits of a mobile device with a projector is that it can project its display onto any surface not normally used with projectors. The surface is not always flat and white. Therefore, an important issue is investigation of adaptive projection techniques based on surface features (e.g., shape, texture, and color) (Raskar et al., 2003) and image contents (Ashdown, Okabe, Sato, & Sato, 2006) to be projected.
- **Real-time calibration of moving projected displays:** In the case where user moves his mobile device and simultaneously conducts manipulations on its projected display, the display should be calibrated as rapidly as possible to make the manipulations easy. Real-time calibration of moving projected displays is useful for applications that augment moving real objects with virtual objects as discussed in Kojima et al. (Kojima, Sugimoto, Nakamura, Tomita, Nii, & Inami, 2006).
- **Precise and rapid finger-gesture recognition:** Recognizing gestures by bare fingers or hands on projected displays is critical, in order to enhance collaboration among users with or without a mobile device. Moreover, from a usability viewpoint, rapid feedback to user manipulations should be provided.
- **Multiuser gesture recognition:** In order to support collaboration adequately among colocated people by using a mobile device with a projector, the device not only should recognize multiple inputs simultaneously given by the people on its projected display but also should identify their individual inputs.
- **New applications for projected displays of mobile devices:** New applications utilizing features of projected displays of mobile devices should be explored. Interesting and inspiring examples are file-transfer methods by overlapping projected displays of mobile devices shown in this chapter, manipulation techniques using another input device such as a laser pointer or a pen (Cao & Balakrishnan, 2006), a system based on a spotlight metaphor (Rapp, Michelitsch, Osen, Williams, Barbisch, Bohan, Valsan, & Emele, 2004), interactions with real objects (Raskar, Beardsley, Baar, Wang, Dietz, Lee, Leigh, & Willwacher, 2004), and so on. Our group has recently developed an entertainment application called COGAME (Interaction Technology Laboratory, 2006). The key idea of COGAME is "manipulation by projection": a player can manipulate a robot by projecting a path through a mobile projector in order to make a robot move on the path.

CONCLUSION

This chapter has described a system called *Hotaru* and possible intuitive manipulation techniques. The design and implementation of *Hotaru* by using currently available technologies were discussed. Informal user studies for evaluating *Hotaru* proved that it would effectively support collaborative tasks in colocated situations. The studies also clarified problems to be solved and examined. We also raised several issues for further research with making a mobile device containing a projector and a camera more useful and usable for our colocated collaboration.

REFERENCES

Ashdown, M., Okabe, T., Sato, I., & Sato, Y. (2006). Robust content-dependent photometric projector compensation. In *Proceedings of 3rd IEEE International Workshop on Projector-Camera Systems* (*ProCams 2006*) (pp. 60–67). New York, NY.

Ayatsuka, Y., Matsushita, N., & Rekimoto, J. (2000). HyperPalette: A hybrid computing environment for small computing device. In *Proceedings of Conference on Human Factors in Computing Systems (ACM CHI2000)* (pp. 133–134). The Hague, The Netherlands.

Cao, X., & Balakrishnan, R. (2006). Interacting with dynamically defined information spaces using a handheld projector and a pen. In *Proceedings of ACM Symposium on User Interface Software and Technology* (*UIST2006*) (pp. 225-234). Montreux, Switzerland.

Holman, D., Vertegaal, R., Altosaar, M., Troje, N. & Johns, D. (2005). PaperWindows: Interaction techniques for digital paper. In *Proceedings of Conference on Human Factors in Computing Systems (ACM CHI2005)* (pp. 591–599). Portland, OR.

Interaction Technology Laboratory. (2006, September 16). *COGAME Project*. Retrieved December, 12, 2006, from http://www.itl.t.u-tokyo.ac.jp/cogame/

Koike, H., Sato, Y., & Kobayashi, Y. (2001). Integrating paper and digital information on EnhancedDesk: A method for real-time finger tracking on augmented desk system. *ACM Transactions on Computer Human Interaction*, *8*(4), 307–322.

Kojima, M., Sugimoto, M., Nakamura, A., Tomita, M., Nii, H., & Inami, M. (2006). Augmented Coliseum: An augmented game environment with small vehicles. In *Proceedings of the First IEEE International Workshop on Horizontal Interactive Human-Computer Systems* (*TABLETOP '06*) (pp. 3–8). Adelaide, Australia.

Lee, J., Dietz, P., Maynes-Aminzade, D., Raskar, R., & Hudson, S. (2004). Automatic projector calibration with embedded light sensors. In *Proceedings of ACM Symposium on User Interface Software and Technology* (*UIST2004*) (pp. 123–126), Santa Fe, NM.

Mitsubishi Electric Pocket Projector (2004). Retrieved December, 2004, from http://www.merl.com/people/raskar/pocket.html

Miyahara, K., Inoue, H., Tsunesada, Y., & Sugimoto, M. (2005). Intuitive manipulation techniques for projected displays of mobile devices. In *Proceedings of Conference on Human Factors in Computing Systems* **(***ACM CHI2005***)** (pp. 1657–1660) Portland, Oregon.

New York Times (2004). *For your viewing pleasure, a projector in your pocket*. Retrieved, November, 2004, from http://www.nytimes.com/2004/11/04/technology/circuits/04next.html

Pingali, G., Pinhanez, C., Levas, A., Kjeldsen, R., Podlaseck, M., Chen, H., & Sukaviriya, N. (2003). Steerable interfaces for pervasive computing spaces. In *Proceedings of the First IEEE International Conference on Pervasive Computing and Communications* (*PerCom2003*) (pp. 315–322). Fort Worth, TX.

Rapp, S., Michelitsch, G., Osen, M., Williams, J., Barbisch, M., Bohan, R., Valsan, Z., & Emele, M. (2004). Spotlight navigation: Interaction with a handheld projection device. In *Proceedings of Pervasive Computing* (*Pervasive2004*). Vienna, (video submission).

Raskar, R., Beardsley, P., Baar, J., Wang, Y., Dietz, P., Lee, J., Leigh, D., & Willwacher, T. (2004). RFIG Lamps: Interacting with a self-describing world via photosensing wireless tags and projectors, *ACM Transactions on Graphics*, *23*(3). 406–415.

Raskar, R., van Baar, J., Beardsley, P., Willwacher, T., Rao, S., & Forlines, C. (2003). iLamps: Geometrically aware and self-configuring projectors. *ACM Transactions on Graphics*, *22*(3). 809–818.

Raskar, R., Welch, G., Cutts, M., Lake, A., Stesin, L., & Fuchs, H. (1998). The office of the future: A unified approach to image-based modeling and spatially immersive displays. In *Proceedings of the 25th Annual Conference on Computer Graphics* (*SIGGRAPH'98*) (pp. 179–188). Orlando, FL.

Rekimoto, J., & Saito, M. (1999). Augmented Surfaces: A spatially continuous work space for hybrid computing environments. In *Proceedings of Conference on Human Factors in Computing Systems* **(***ACM CHI99*) (pp. 378–385). Pittsburgh, PA.

Roeber, H., Bacus, J., & Tomasi, C. (2003). Typing in thin air: the Canesta projection keyboard—A new method of interaction with electronic devices. In *Proceedings of Conference on Human Factors in Computing Systems (ACM CHI2003)* (pp. 712–713). Fort Lauderdale, FL.

Sugimoto, M., Miyahara, K., Inoue, H., & Tsunesada, Y.(2005). Hotaru: Intuitive manipulation techniques for projected displays of mobile devices. In *Proceedings of International Conference on Human-Computer Interaction (INTERACT2005)* (pp. 57–68). Rome.

Sukthankar, R., Stockton, R., & Mullin, M. (2001). Smarter Presentations: Exploiting homography in camera-projector systems. In *Proceedings of International Conference on Computer Vision* (*ICCV2001*) (pp. 247–253). Vancouver, B.C., Canada.

Summet, J., & Sukthankar, R. (2005). Tracking locations of moving hand-held displays using projected light. In *Proceedings of Pervasive Computing* (*Pervasive2005*) (pp. 37–46). Munich, Germany.

Telecommunication Carrier Association (1996). Retrieved July 2006, from http://www.tca.or.jp/eng/database/daisu/yymm/0607matu.html

Trucco, E., & Verri, A. (1998). *Introductory techniques for 3-D computer vision*, Upper Saddle River, NJ: Prentice Hall.

Welch, G., & Bishop, G. (2001). An introduction to the Kalman filter, *SIGGRAPH 2001 Tutorial*, Retrieved November, 2004, from http://www.cs.unc.edu/~tracker/media/pdf/SIGGRAPH2001_CoursePack_08.pdf

Wellner, P. (1993). Interacting with paper on the DigitalDesk. *Communications of the ACM*, *36*(7), 87–96.

Zhang, Z., Wu, Y., Shan, Y., & Shafer, S. (2001). Visual Panel: Virtual mouse, keyboard and 3-D controller with an ordinary piece of paper. In *Proceedings of Perceptual User Interface* (*PUI2001*) (pp. 1–8). Orlando, FL.

KEY TERMS

Colocated Collaboration: Colocated collaboration is a type of face-to-face collaboration where people share their place and time. However, colocated collaboration is different from a traditional face-to-face collaboration in that the former is usually discussed in the context of a computationally augmented environment.

Finger Gesture Recognition: Finger gesture recognition is the interpretation of human finger gestures and the utilization of this as input for a computer. Computer vision techniques or those with inertial sensors (e.g., accelerometers) are mainly used for the recognition.

Intuitive Manipulation Technique: Using a mouse and a keyboard as an input device for a computer is often irritating and awkward. Although less awkward and bothersome techniques using a mouse and keyboard have been investigated, one of the intuitive manipulation techniques is to use different devices or methods, for example, natural human gestures such as finger/hand gestures, speech/sounds and so on, as input to a computer.

Location-Aware Projection: Location-aware projection is the projection of a graphical image generated by a computer onto a surface based on the 3-D position and orientation of the computer or the locations of people who view the image.

Manipulation by Projection: Manipulation by projection is a new technique that allows a user to manipulate an object by overlaying an image using a mobile projector. For example, by projecting a map image on an area where a robot is placed, a user can make the robot follow a path on the map. A user can translate or rotate the projected map by changing his standing point or rotating his mobile projector, in order to make the robot move as he intends.

Movable Projected Display: A projected display is a screen image of a computer projected onto a surface through a projector and used as a display for the computer. A projected display becomes movable when a computer with a projector is light and small, such as a mobile device.

Projectable Interface: A graphical user interface of a computer usually appears on its display and is manipulatable by using a mouse or a keyboard. A projectable interface is a user interface projected onto any surface such as a wall, and manipulatable by pointing or touching it with a laser pointer or a finger.

Section III
Assistive Mobile Technologies

The potential for well-designed mobile technologies to fulfill an assistive role—that is, as assistive technologies—is increasingly being realised. This section presents a selection of chapters that discuss the design implications specific to assistive mobile technologies. The section starts with a discussion of the key issues and trends of designing and evaluating mobile assistive technologies before looking at design to support specific demographics, including seniors, mental health interventions, and visually impaired users. The chapter concludes with a look at designing the user interface to the ultimate assistive mobile device, the smart wheelchair.

Chapter XXXVI
Designing Mobile Technologies for Individuals with Disabilities

Rock Leung
University of British Columbia, Canada

Joanna Lumsden
National Research Council of Canada, Canada

ABSTRACT

While mobile devices offer many innovative possibilities to help increase the standard of living for individuals with disabilities and other special needs, the process of developing assistive technology, such that it will be effective across a group of individuals with a particular disability, can be extremely challenging. This chapter discusses key issues and trends related to designing and evaluating mobile assistive technology for individuals with disabilities. Following an overview of general design process issues, we argue (based on current research trends) that individuals with disabilities and domain experts be involved throughout the development process. While this, in itself, presents its own set of challenges, many strategies have successfully been used to overcome the difficulties and maximize the contributions of users and experts alike. Guidelines based on these strategies are discussed and are illustrated with real examples from one of our active research projects.

INTRODUCTION

In an assistive capacity, computer technology can play an important role in helping to increase the standard of living for individuals with, for example, physical, cognitive, developmental, psychiatric, and learning disabilities. Technology can also benefit individuals with other special needs, for example, users with limited literacy skills and older users (age 65+). Technology in this genre (commonly known as *assistive technology*) can play a rehabilitative role to help individuals overcome a disability; it can also help individuals with a disability or special need to perform a particular activity associated with daily living that they would otherwise be unable to do, thus providing a degree of independence. For the purpose of discussion in this chapter, we refer collectively to persons with disabilities and those with other special needs as *individuals with disabilities.*

Mobile computer technologies have many features that make them well suited to delivering assistive support. Being portable, they can be used to assist in various settings beyond the desktop. Recent mobile devices also have the capacity to exploit location and context information, and are thereby able to provide more advanced and/or intelligent assistance. Mobile devices are also becoming increasingly more powerful in terms of computing power, memory storage, and network capabilities. The cost of mainstream mobile computing is, however, relatively inexpensive compared to the cost of traditional desktop computers, which makes mobile devices more financially accessible.

Mobile devices can help individuals with disabilities in a number of ways (Bertini & Kimani, 2003) by, for example, acting as:

- An aid to carry out daily activities (e.g., individuals with a physical disability can use a device to remotely issue commands to operate PCs, elevators, doors, etc.)
- A means to communicate with others
- A guide or advisor that exploits contextual information to proactively help or warn the individuals

Despite the relative infancy of mobile technologies per se, many assistive applications have already been researched and developed based on commercial handheld devices and/or mobile phones. For example: Myers et al. (Myers, Wobbrock, Yang, Yeung, Nichols, & Miller, 2002) developed software to run on Palm OS and other mobile devices that assists individuals with neuromuscular disorders (e.g., Cerebral Palsy) to enter text into a computer; Fischer and Sullivan (2002) designed a proof-of-concept system for reminding users with cognitive disabilities through location-aware mobile phones about when to get on and off public transportation; Davies et al. (Davies, Stock, & Wehmeyer, 2002) have developed multimedia prompting software on a Windows CE mobile device to help users with mental retardation to complete work tasks; Wu et al. (Wu, Baecker, & Richards, 2005) created a software application to run on a Palm OS device that helps prevent users with amnesia from experiencing disorientation; and Moffatt et al. (Moffatt, McGrenere, Purves, & Klawe, 2004) developed a sound- and image-enhanced schedule planner to run on a Windows CE device to help users with aphasia. To our knowledge, all of these applications, except the proof-of-concept prototype by Fischer and Sullivan (2002), have been implemented as fully functional prototypes that have been tested in the field by target users.

While mobile devices offer many possibilities for innovative assistive technologies, the process of developing an assistive technology, such that it will be effective across a group of individuals with a particular disability, can be extremely challenging. First, designing assistive mobile technology must consider whether inherent properties of a mobile device platform may make using mobile applications more difficult for certain users with disabilities (e.g., small screen size for people with visual impairments, small physical controls for people with neuromuscular disorders). Further, each user population has unique needs and abilities that heavily influence both the design and the design process; furthermore, individual user characteristics often vary considerably across the members of a user population and, if suffering from a degenerative disability, any one individual's needs may change on a day-to-day basis. In a group of persons with vision impairment, for example, individuals may differ greatly in terms of their level of impairment, as well as other usability-related factors such as age, experience with computers, and physical capabilities. Although the same might be said of any user population, given the very specific needs and abilities of individuals with disabilities, researchers and/or developers often find it hard to relate to and understand the requirements and/or constraints of such populations and their individual members; hence, the design process is often considerably more challenging than the norm.

Given the infancy of the field of mobile assistive technologies, knowledge regarding the effectiveness and applicability of design and evaluation methods for use in this field is currently evolving. This chapter discusses key issues and trends related to designing and evaluating mobile assistive technology for individuals with disabilities. Following an overview of general design process issues, we argue (based on current research trends) that individuals with disabilities (i.e., the target users) and domain experts (e.g., clinicians, caregivers, academics) be involved throughout the development process. While this, in itself, presents its own set of challenges, many strategies have successfully been used to overcome the difficulties and maximize the contributions of users and experts alike. Guidelines based on these strategies are discussed and are illustrated with real examples from one of our active research projects.

BACKGROUND

Designing *any* technology requires a good understanding of the target users' needs and abilities. This requirement is particularly crucial yet complicated when designing assistive technologies for individuals with disabilities. When designing such technology, *all* aspects of users' cognitive, physical, and sensory capabilities need to be taken into account, not just those directly related to the disability (LoPresti & Willkomm, 1997); for example, people with cognitive disabilities often have additional physical and sensory limitations (e.g., difficulty with vision, hearing, tactile sense, fine motor control, ability to speak, coordination) (LoPresti, Mihailidis, & Kirsch, 2004). In addition, specific emotional and behavioural changes associated with a disorder, for example, may influence a user's motivation to use (and sustain effort over time with) an assistive technology. Furthermore, users' attitudes towards technology, and assistance/treatment (especially those that appear to exert "external control"), also need to be factored into the design process because such attitudes often influence users' acceptance of proposed assistive technologies (LoPresti et al., 2004). Understanding (or relating to) these needs is often difficult for researchers and developers in technical fields, who typically have limited knowledge about, and experience with, the disability.

Recognising the limits of their understanding, many researchers have involved target users in the design and evaluation process for assistive technologies; as a result, they have typically obtained valuable results. Assistive technology development projects regularly involve target users in prototype evaluation. Such users often participate in experiments (e.g., Davies et al., 2002; Stevens & Edwards, 1996; Tee, Moffatt, Findlater, MacGregor, McGrenere, Purves, & Fels, 2005) to identify usability issues, and to see how well the technology assists the user. A variety of other mainstream evaluation techniques, such as interviewing (e.g., Davies et al., 2002; Lumsden, Leung, & Fritz, 2005; Tee et al., 2005), video analysis of users' interactions with technology (e.g., Carmien, 2005; Leung, Lumsden, & Fritz, 2006), workload assessment (e.g., Leung et al., 2006; Stevens & Edwards, 1996), and field studies (e.g., Fischer & Sullivan, 2002; Wu et al., 2005), have also been used. Recently, a number of research projects have reported involving target users in the design process itself, often using participatory design methods (e.g., Lumsden et al., 2005; Moffatt et al., 2004; Wu, Richards, & Baecker, 2004). In light of the benefits and past successes, it is strongly recommended (Bergman & Johnson, 1995; LoPresti et al., 2004) that target users be involved in all stages of the development of assistive technologies.

In addition to involving target users, many research projects have involved domain experts, to compensate for the limited knowledge and experience of the research team, in a variety of roles (Allen, Leung, McGrenere, & Purves, in submission), including:

- *As a member of the research/planning team* where, in order to inform the research/design process, the domain expert provides expertise on the disability and working with individuals with the disability (e.g., Davies, Marcella, McGrenere, & Purves, 2004; Mihailidis & Fernie, 2002; Moffatt et al., 2004; Sullivan & Fischer, 2003; Tee et al., 2005; Wu et al., 2005)

- *As a liaison between the research/design team and target users* where the domain expert can help the users gain trust in the research process, facilitate communication if needed, and recruit participants (e.g., Leung et al., 2006; Lumsden et al., 2005; Moffatt et al., 2004; Tee et al., 2005)
- *As a representative* where the domain expert participates as a representative for one or more target users in the research/design process (e.g., Lumsden et al., 2005; Moffatt et al., 2004)

Domain experts clearly have a great deal to contribute, and many have strongly recommended their involvement in the research and design of assistive technology (LoPresti et al., 2004; Sullivan & Fischer, 2003; Tee et al., 2005; Wu et al., 2004).

General acceptance that target users and domain experts be involved in the research and design of assistive technology does not, however, mean that involving members of these two stakeholder groups in the design process is a simple or straightforward undertaking; the associated challenges are outlined below.

Challenges of Working with Target Users

A frequently reported issue associated with involving target users in the design of assistive technologies is finding and recruiting enough representative participants (e.g., LoPresti et al., 2004; Moffatt et al., 2004; Stevens & Edwards, 1996; Tee et al., 2005). Some research methods, such as experimental evaluations or surveys, typically require a large sample of participants in order to return statistically valid results. The available population of individuals with a particular disability is, however, often relatively small; as such, it can be challenging to find potential participants to recruit. Furthermore, the lack of homogeneity of disability, and other relevant factors across target users within any one population, makes it difficult to determine selection criteria that would lead to a representative, yet manageable, user group for inclusion in a study.

Even when potential participants *are* identified, obtaining informed consent to work with target users may be difficult (Newell & Gregor, 2000); disabilities that affect an individual's cognitive or communicative abilities may prevent a user from understanding the nature of his/her involvement and, in turn, make it difficult for a researcher to ensure that the individual is indeed giving *informed* consent. Cognitive and communicative disabilities, in particular, can also make it difficult to work with individuals throughout their involvement in the research. As previously mentioned, domain experts often, therefore, serve as liaisons to facilitate communication with target users during the users' involvement in a project.

Although target users *have* been involved in experiments to evaluate assistive technologies, it is often difficult (if not impossible) to impose the level of experimental control that is typical of quantitative experiments. Having a common testing site (e.g., a controlled lab), for instance, may be unsuitable because of the nature of the target users' disability, target users may have difficulty travelling to a particular testing site, and/or such users may be uncomfortable being in an unfamiliar environment. Tee et al. (2005), for example, reported that while they would have preferred that their evaluation take place at a single location, 50% of the participants required the researchers to come to their homes (for reasons of mobility and comfort), while the other 50% were not willing to have researchers in their homes. Aside from meeting in different locations, it is often very difficult to ensure that different groups of participants are similar, and that they receive identical treatment (Stevens & Edwards, 1996). Additionally, when the assistive mobile technology being evaluated is *truly* innovative, it is impossible to introduce an existing alternative technology as a control or base line against which to accurately assess its performance (Stevens & Edwards, 1996).

As previously indicated, one of the significant advantages of mobile technologies in terms of their applicability to assistive support functions is their portability; they have the potential to play a supportive role in many different contexts of use. This, however, poses a significant challenge in terms of evaluating the assistive devices; although effective evaluations of mobile devices require the evaluation design to meaningfully reflect the intended context of use (Lumsden, Kondratova, & Langton, 2006), evaluating mobile assistive technology with target users may be extremely challenging for some anticipated use-case scenarios and contexts. Ethically, one has to be particularly conscious of the potential consequences of evaluating a mobile assistive device under real-use conditions in terms of the level of risk to the participants and the degree of harm associated with device failure; both these issues being heightened as a result of the participants' disabilities.

Challenges of Working with Domain Experts

Working with domain experts is not, however, without its own set of challenges. Difficulties may, for example, arise over time when researchers and domain experts hold mismatched expectations of project goals (Allen et al., in submission). Problems (often unexpected) may also occur when domain experts assume more than one of the three roles identified previously, and the responsibilities of each role are not clearly defined or understood (Allen et al., in submission). Allen et al. (in submission), for example, describe a project that utilized two domain experts to act as representative users in the design phase of its study and then act, in the evaluation phase, as liaisons to people under their care: by the end of the design phase, the domain experts developed an unstated expectation that the design would be developed into a working product that target users in the evaluation phase would be able to keep and use long term; once they realized that only a research prototype was going to be developed for the evaluation, they felt they could not recruit their clients for the study.

Complications can occur when domain experts act as a liaison to a community of individuals with a disability, but inadvertently do not communicate information (e.g., research goals, the nature of participant involvement) accurately to these individuals. Domain experts working with target users, especially users with disabilities affecting communication, often need to communicate in a form that the users can comprehend; some information can occasionally be lost or altered in the "translation."

GUIDELINES FOR INCLUSIVE DESIGN OF ASSISTIVE TECHNOLOGIES

Various guidelines for successfully involving, and thereby maximizing the contribution of, target users and domain experts in the design and evaluation of assistive mobile technology are currently evolving. Each of these guidelines is discussed in detail.

Work with Existing Support Organizations

Working closely with existing organizations that support individuals with a particular disability is often extremely beneficial (Moffatt et al., 2004; Wu et al., 2005). Most importantly, by working with such support organizations, it is typically easier to find and recruit target users for involvement in a research and/or development project. Furthermore, support organizations are typically a solid source of domain experts, paid employees (e.g., clinicians, caregivers) and volunteers who have considerable expertise related to the disability and experience working with individuals with the disability, who may be interested in becoming involved in a research/development project. Involving support organizations in a research and/or development process may also help ensure that those they support are treated appropriately and ethically throughout the project. Additionally, support organizations often allow a research team to conduct research and development activities within the organization's facilities; individuals supported by the organization are normally familiar with the facilities and as such are more at ease in the familiar environment. Finally, special transport arrangements are (where necessary) normally in place for clients attending support organizations and, as such, research teams do not need to worry about appropriate means to transport these individuals. One drawback, however, of using an organization's facilities is the reduced level of control the development team can exert over the study environment; this may lead to unavoidable disturbances throughout the course of the research (Moffatt et al., 2004; Wu et al., 2005).

Assess Target Users' and Domain Experts' Needs, Abilities, and Expectations

Given the aforementioned scarcity of accessible individuals with disabilities and/or domain experts, it is tempting for researchers or developers to include target users or potential collaborators without assessing their needs, abilities, and expectations, simply because they have expressed interest in the project. Gaining a good understanding of these personal characteristics of potential users and experts may, however, avert future problems in the research process or, at the very least, help increase the productivity of time spent with these participants.

As already mentioned, a thorough understanding of all aspects of a user's cognitive, physical, and sensory capabilities, as well as their attitudes towards and associated experiences of technology and assistance/treatment, is valuable in many ways. Although the involvement of target users at all stages of development is recommended, a user's contribution in the process "varies with their skills, experience, and is also dependent on the particular phase of the research or development" (LoPresti et al., 2004, pg. 40). A thorough understanding of target users will, therefore, help to determine appropriate levels of participation. Wu et al. (2004) recommend that, if target user participants will be required to interact with each other during the project, researchers should try to understand how the users' disabilities influence such interaction. They also recommend that researchers should gain an understanding of whether the target users are aware of their own limits, and what strategies are being used to compensate for the disability (Wu et al., 2004).

Standardized testing of target user's disability-related abilities has been found to be useful. Moffatt et al. (2004), for example, noted that standardized testing provided valuable insights into their evaluation results that would not otherwise have been apparent. In addition to standardized testing, ethnographic studies can help researchers understand each target user, as well as identify potential problem activities in context (Wu et al., 2004).

A thorough assessment of domain experts' abilities and expectations is also valuable in fostering good working relationships during the course of a research/development project. Not only is it important to understand a domain expert's level of expertise and experience with target users, it is also important to understand that individual's interest in the research, particularly in terms of his/her perspectives and expectations (Allen et al., in submission). This is crucial in working through the previously noted problems that any mismatched expectations may cause.

Choose a Design/Evaluation Technique and Analyze its Requirements

Only after thorough assessments have helped to gain an understanding of target users' needs and abilities is it possible to choose an appropriate design/evaluation technique to adapt to these users. Many "mainstream" design/evaluation methods have been successfully used with individuals with disabilities to design assistive mobile devices (see Background section). When choosing a technique, Wu et al. (2004) recommend asking: What are the goals of the technique?; How many people are involved?; What is the participation model?; and What phases or activities are involved in the technique?

After a technique has been chosen, it is important to analyze its demands on the target user (Wu et al., 2004) by assessing the cognitive and physical requirements of the technique and how they are related to the technique's goals, structure, and participation model. Some aspects of these requirements are fundamental to the activity (e.g., a group discussion requires communication), while other aspects are flexible (e.g., the communication can be oral or gestural, or involve a liaison to facilitate). Identifying the flexibility in the technique helps to determine how the technique can be suitably adapted to individuals with disabilities.

Adapt the Chosen Approach to be Sympathetic to the Target Users' Abilities

Once selected, a technique can be adapted as appropriate in light of its identified requirements and the target users' needs and abilities. This is a creative, and often iterative, process since the technique may need to be refined after reflecting on lessons learned from initial attempts to use it. It may be useful to ask domain experts for feedback on the suitability of the adapted technique for the intended target users.

Wu et al. (2004) suggest a number of questions that may help in effectively adapting a chosen technique:

- How can human functions be supported using technology or other nontechnological practices?
- Can impractical activities in the technique be avoided or changed to capitalize on the strengths of the target users?
- Are there features from other techniques that can be useful here and if so, is it possible to integrate those features into this technique?

It may not always be possible to adapt a technique for target users and then obtain the same type of research data as is typically returned by the technique in its original form. In the case of experimental evaluations involving individuals with disabilities, researchers have often had to loosen some requirements typical of a controlled experiment, and adapt their experimental design according to the target users' unique needs (e.g., Moffatt et al., 2004; Stevens & Edwards, 1996; Tee et al., 2005). As previously noted, it may not be possible to conduct the experiments at the same location for all participating target users; it may also be difficult to conduct the evaluation exactly the same way with all target users, especially if the users have difficulty communicating, and small sample sizes may make it difficult to balance the characteristics of participants between research groups. Fundamentally, if it is only possible to recruit a small sample of target users, evaluations may have to be based more on case studies and qualitative results than statistical analysis and quantitative results. Despite the fact that it can prove difficult to control many experimental variables when involving individuals with disabilities, past experiments have been successful at obtaining valuable insights both about the target user population, and for improving the usability of the evaluation prototype.

Clearly Communicate the Nature of Participants' Involvement

After structuring a research or development protocol to involve target users and domain experts, it is essential to effectively communicate to them the precise nature of their involvement. While issues such as mismatched expectations can easily manifest and harm a project, they can be as easily avoided through clear communication at the start of the participants' involvement. When communicating via a domain expert with target users whose communication skills are impaired, it is important to ensure that the message is accurately interpreted and delivered to the target users. It is particularly critical to effectively communicate the project's goals, especially if the users and experts have never before been involved in a research or development project. Newell and Gregor (2000), for example, strongly emphasise the importance of helping target users and domain experts understand how the motivation and methodologies of the research and development process(es) are different from clinical situations with which they are likely more familiar: it is important for target users to understand that some research may return negative results, which may disappoint the participants; it is also important, at the start of the research, to make clear to both target users and domain experts that the provision of any long-term support is not the responsibility of the project.

Attempt and Refine the Approach

After a newly adapted technique has been deployed, it is important to watch for potential negative consequences and evaluate its effectiveness (Wu et al., 2004). It is then important to revise and improve the technique, as needed, based on what was learned in practice. The experience gained in working with target users is also extremely valuable for learning how to better interact with them, often helping the research team ensure that the project goals are met. Moffat et al. (2004), for example, documented a project where the most difficult challenge was communicating effectively with target users:

Extra time had to be allocated to ensure participants had sufficient time to fully understand the tasks and ask questions. While this was a challenge throughout all phases of this research, it was particularly significant during the experimental evaluation phase, where timing was critical. In that state, the sensitivity developed by the researchers during the [previous] participatory design phase was critical to minimizing the effect of communication barriers on the research outcome. (Moffatt et al., 2004, p. 414).

Evaluate the Technology in Different Contexts

Assistive mobile technology should be evaluated in the many contexts (e.g., at home, at the support organization, outdoors) in which target users spend their lives (LoPresti et al., 2004). Mobile technology is often evaluated in a controlled setting (e.g., laboratory), which may be very different from the environments in which individuals with disabilities live. Controlled evaluations are valuable for initial usability assessments of technologies, especially when representative contexts of use are incorporated into the lab-based protocol (e.g., Lumsden et al., 2006); in contrast, evaluating the use of the assistive technology in the "real world" often helps the research or development team better understand whether the technology will be usable and effective over the long term. Field trials (e.g., Wu et al., 2005) and ethnographic studies (e.g., Davies et al., 2004) have been conducted to gain this type of understanding.

GUIDELINES IN USE: REFLECTIONS ON A CASE STUDY

To illustrate practical application of these guidelines, the remainder of this chapter describes a case study in which we applied the guidelines during the design and evaluation of a specific mobile assistive technology. The case study on which we will focus is the ALEX project, the aim of which is to design a mobile *A*dult *L*iteracy support application for *EX*periential learning—that is, a handheld application to assist adults with limited literacy skills in their daily lives. ALEX is designed with speech recognition and synthesis capabilities to facilitate, in a manner sympathetic to the needs of functionally illiterate adults, the transcription of spoken words to text, the editing of text, and the use of a series of language-related support tools, such as a dictionary (for more detailed information on this work see (Leung et al., 2006; Lumsden et al., 2005)).

Work with Existing Support Organizations

At the onset of the research project, we contacted several local literacy organizations to engage their involvement in the project. As a result, we have subsequently worked with four literacy organizations that expressed interest and allowed us to recruit interested literacy facilitators/tutors (domain experts) and literacy students (target users) to participate in our research. Three of the literacy organizations also allowed us to conduct aspects of the research at their facilities. Despite the fact that nearly 25% of adults in the world's richest countries are reported to be functionally illiterate (UNSECO, 2000), adult illiteracy retains an element of social stigma that often means adults who are functionally illiterate are very private about their "social disability." Literacy organizations serve those adults who have acknowledged the need, and are actively seeking to improve their literacy levels. By working directly with existing literacy organizations, we were granted access to a population that would have otherwise been very hard to identify and recruit. Furthermore, many adults with limited literacy skills prefer to remain within known comfort zones (Lumsden et al., 2005) and so by being introduced to the target users via a vehicle with which they were familiar, as well as being able to conduct some of the research in an environment familiar to the target users, we feel we were better able to put our participants at ease and thereby maximize their involvement and contribution to the project.

Assess Target Users' and Domain Experts' Needs, Abilities, and Expectations

We attributed considerable time and energy to assessing our participants' needs, abilities, and expectations not just at the beginning of our research, but in an ongoing capacity throughout the work. When we first approached both the target users and domain experts in order to engage them in our project, we asked them informally about their thoughts on the stated project goals. Additionally, we asked the domain experts to reflect on the number of years of experience they had accumulated in supporting literacy students.

After recruiting six adult literacy students and three literacy facilitators to our research project, we conducted a series of focus groups (informal semistructured small group discussions) in order to better understand various aspects of the literacy students' lives, such as their educational background, how they cope with their literacy

Table 1. Participant characteristics: Literacy students

Participant ID	Estimated Age	Gender	Literacy Level	Experience with Computers
S1	late teens	female	In high school and can read and write, but behind in her classes	Moderate
S2	mid 20s	male	Joined literacy organization to improve math skills, but may have limited literacy skills as well	Moderate
S3	late 30s	female	Can read and write; studying to obtain GED diploma to get better job	Little
S4	mid 40s	female	Almost no literacy skills; only 2 years of formal education	Little
S5	mid 40s	male	Can read and write; studying to obtain GED diploma	Moderate
S6	late 40s	male	Can read and write; studying to obtain GED diploma	Almost none

Note: GED = General Educational Development; a GED diploma is considered equivalent to a High School Graduation Diploma (Government of New Brunswick, Department of Post-secondary Education, Training and Labour Apprenticeship and Certification, 2003) deficits, and their experience with and views on mobile technology. Tables 1 and 2 summarize some participant characteristics gathered from the focus groups. Through these focus groups, we learned that the literacy students tended to prefer to operate within preestablished comfort zones, and employed a variety of coping strategies to make up for their limited literacy skills. We also learned that the literacy students had a range of experience using technology, but had little experience using mobile devices. Furthermore, we learned that the students were interested in using handheld computers and other mobile devices to assist them in their daily lives, in particular because they viewed such technology as a "status leveller."

Having elicited an extensive amount of valuable information about the participants' requirements and associated abilities, we invited a subset of the focus group participants (based on their level of active participation up to that point) to take part in the actual design process for the assistive technology itself. Four literacy students (S3, S4, S5, and S6) and one facilitator (F2) agreed to participate. Throughout this process (which is discussed further) we continued to elicit information about participants' expectations and needs in relation to their abilities.

Our experience cemented in our minds the importance of placing the anticipated end users of assistive technology in a central, inclusive role in the design of the technology. Without such close involvement of members of our user group, we would not have been able to relate to the problems, preferences, and coping strategies of our target users. When one does not suffer from a particular "disability," it is extremely hard to anticipate the complexities (and perhaps, to an even greater extent, the knock-on complications) associated with living with the condition. It is only through the direct involvement of our target users and domain experts that we were able to effectively assess their needs and expectations. This knowledge, we hope, makes us far better able to deliver an assistive application that is designed to be sympathetic to their abilities whilst meeting their vision.

Choose a Design/Evaluation Technique and Analyze its Requirements

Although we have used a variety of design techniques at various stages of the project to date, in the interests of brevity, we will focus here on the technique we used for the active design process. Continuing our commitment to place our target users in a central role, we selected to use a participatory design (PD) approach, principally because this type of approach generally involves target users as full and *equal* participants in the design process. Many design and evaluation techniques involve considerable use of written forms of information delivery and communication. This was obviously wholly inappropriate for our target users: we therefore chose to use the PICTIVE (*P*lastic *I*nterface for *C*ollaborative *T*echnology *I*nitiatives through *V*ideo *E*xploration) participatory design method (Muller, 1992). PICTIVE is a semiformalized PD approach that is used to produce a prototype from common office supplies rather than text documents or computer software. These familiar office supply materials allow everyone on the design team to contribute equally to the final paper prototype, regardless of level of literacy (or indeed technological savvy). PICTIVE was also chosen because it had

Table 2. Participant characteristics: Literacy facilitators

Participant ID	Gender	Experience as a Literacy Facilitator	Experience with Computers
F1	male	6 months	Almost none
F2	female	1 year	Limited
F3	female	5+ years	Moderate

been found, on many past commercial software development projects where it was used, to be very enjoyable by all design team members. As researchers, we wanted the participants to *enjoy* their design experience: taking part in design work was a new (and potentially intimidating) task for the literacy students and we wanted them to feel relaxed, able to participate, and empowered. We felt strongly that the PICTIVE PD method was ideal for our purpose.

Adapt the Chosen Approach to be Sympathetic to the Target Users' Abilities

Although we felt that, in essence, the PICTIVE PD method was ideally suitable as a means to involve our target users in the design process, we adapted it in terms of making some additional allowances to accommodate the needs of our particular participants. Specifically, we were sensitive to the fact that the whole process of designing software was outside the comfort zone of our participants, and so we made a concerted effort to regularly remind participants that they were *part* of the design team and that we considered them the *experts* on their needs and daily contexts. Additionally, we continuously validated their ideas to provide ongoing encouragement, and we relied heavily on images and graphics when communicating concepts and ideas to participants. The design team was deliberately assembled with many more participants than researchers (5 to 1 instead of an equal number of each), in the hope that the imbalance in numbers in their favour would help them feel more comfortable. As already mentioned, these necessary adjustments were only highlighted as a result of the effort we expended in assessing our target users prior to their direct involvement.

Clearly Communicate the Nature of Participants' Involvement

At all points in this research project, we have made a deliberate effort to ensure that participants understood what was asked of them, and to explain the project goals. All textual information (principally, consent forms) was written in a language that we felt was appropriate for the students' literacy levels in an attempt to make it easier to understand; images and graphics were used to supplement the text. In addition, any printed information was presented verbally (in person) to the literacy students, and they were given an opportunity to ask questions. Information was communicated to our domain experts (the literacy facilitators) in text and via personal presentations to ensure that their involvement was clear and to enforce the goals of the project. As a result of the care and attention we have given to communication throughout the process, to date, we have not encountered any obvious signs of mismatched expectations or lack of understanding in terms of the role of our various participants.

Attempt and Refine the Approach

While the participatory design process (using the PICTIVE approach) proceeded smoothly, we did make some minor changes to maximize the design team's productivity: in particular, we modified the duration of design sessions and the total number of sessions completed. When we were initially finalizing our protocol, we thought that four 2-hour long design sessions would be appropriate and adequate. At the end of the initial design session, however, it became clear that 2 hours was too long for participants to sustain productive design work. In addition, we encountered difficulties when trying to schedule all participants to meet together for the remaining three sessions. In response to these observations, we decided to complete six 60-90 minute long design sessions, and we asked participants to come to as many sessions as they could (but we did not expect them to come to all of the sessions). This new schedule appeared to have worked well and the design work was completed (see Figure 1 for images of the designed prototype).

After completing our design sessions, we asked each participant to reflect on their experience as members of the design team. They were each asked to react to a series of questions using a

Figure 1. a) Prototype design and b) Prototype in a target user walkthrough

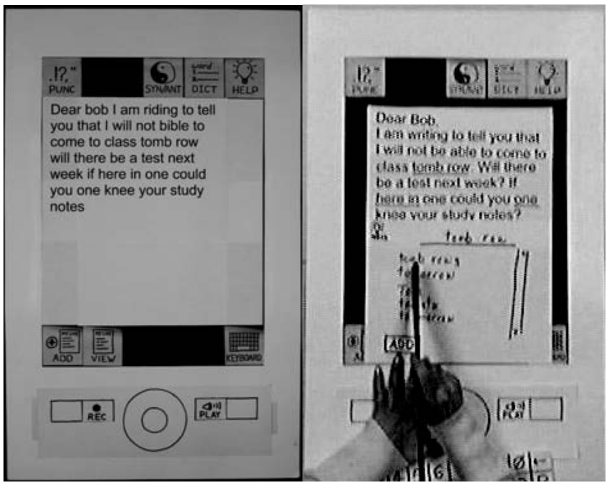


Table 3. Average responses to post-design session questions

Question	Individual Responses				Average Response
Did you enjoy the design sessions?	5	4	4	5	4.5
How understandable was the aim of the design meetings?	4	4	5	5	4.5
How easy were the steps in the design process to understand?	4	3	5	4	4.0
How easy was it to understand what a handheld computer could and could not do?	4	4	3.5	1	3.2
How easy was it to come up with scenarios?	5	4	5	5	4.75
How easy was it to contribute to the paper prototype design?	5	4	5	5	4.75
How important do you feel your contributions were to the design?	4	5	5	5	4.75
How useful were the office supplies in terms of letting you show the team your ideas?	5	5	5	5	5.0
Was your time in the meeting sessions well spent?	5	5	5	5	5.0
What do you think about the quality of the design?	5	5	5	5	5.0
Do you think the final design meets the needs of the literacy students in the team?	5	5	5	5	5.0

Note: Individual responses were collected anonymously.

five point scale (with 1 = lowest score; 5 = highest score). The questions and responses are shown in Table 3. Participants viewed their experience and the value of the resulting prototype design very highly. Perhaps unsurprisingly, for many participants, the most challenging aspect of the design process was understanding the capabilities of mobile technology; this did not, however, prevent the participants from actively engaging in the design process.

Our observations, together with the resulting prototype design and the participant feedback, indicated the PD sessions were a positive, successful experience for all participants. We felt that each member provided valuable contributions, irrespective of his or her particular literacy levels. We also observed much enthusiastic and synergistic collaboration by the design team members throughout the process. For example, at a particular session when some design decisions were being made, participants with higher literacy levels expressed much interest in knowing what the participant with very low literacy levels (S4; who on that occasion was absent) thought of the decisions, and whether she would be able to use the revised design. At the end of the final design session, each of the participants expressed disappointment; every one of them would have happily continued with the process. Our tailored PICTIVE participatory design method was, certainly in this instance, a valid and valuable tool for design activities involving participants with limited literacy skills.

Evaluate the Technology in Different Contexts

To date, implementation of the ALEX system as a high-fidelity prototype that we can use for evaluations across multiple contexts has not yet been completed. It is, however, our intention to conduct extensive evaluations of the prototype across multiple contexts in both the lab and field in order to determine its usability and the longitudinal impact the system has on the literacy levels of its users.

CONCLUSION

In this chapter we have highlighted the challenges associated with developing mobile assistive technologies for individuals with disabilities. From research published in this area, we have drawn a set of guidelines intended to help overcome the challenges inherent in involving target users with disabilities and domain experts in the design and evaluation process for such technology. As our reflections on our own case study illustrate, with careful thought and adaptability to the needs, abilities, and expectations of the specific target user population, it is possible to successfully integrate individuals with disabilities and domain experts into the design process for innovative mobile assistive technology. Research in this field is in its infancy; the overview we present here serves merely as a starting point to collate experience and guidelines to help improve and support the design and evaluation of mobile assistive devices.

REFERENCES

Allen, M., Leung, R., McGrenere, J., & Purves, B. (in submission). *Involving domain experts in assistive technology research*. University of British Columbia.

Bergman, E., & Johnson, E. (1995). Towards accessible human-computer interaction. In J. Nielsen (Ed.), *Advances in human-computer interaction*, (vol. 5, pp. 87-113). Norwood, NJ: Ablex Publishing.

Bertini, E., & Kimani, S. (2003, September, 8-11). Mobile devices: Opportunities for users with special needs. In *5th International Symposium on Mobile Human Computer Interaction (Mobile HCI'2003)* (pp. 486-491). Udine, Italy.

Carmien, S. (2005, April 2-7.). End user programming and context responsiveness in handheld prompting systems for persons with cognitive disabilities and caregivers. In *Human Factors in Computing Systems—Extended Abstracts (CHI'2005)* (pp. 1252-1255). Portland.

Davies, D. K., Stock, S. E., & Wehmeyer, M. L. (2002). Enhancing independent task performance for individuals with mental retardation through use of a handheld self-directed visual and audio prompting system. *Education and Training in Mental Retardation and Developmental Disabilities, 37*(2), 209-218.

Davies, R., Marcella, S., McGrenere, J., & Purves, B. (2004). The ethnographically informed participatory design of a pd application to support communication. In *6th international ACM SIGACCESS Conference on Computers and Accessibility (ASSETS'2004)* (pp. 153-160). Atlanta, USA. October 18-20.

Fischer, G., & Sullivan, J. F., Jr. (2002, June). Human-centered public transportation systems for persons with cognitive disabilities—Challenges and insights for participatory design. In *Participatory Design Conference (PDC'02)* (pp. 194-198). Malmö University, Sweden.

Government of New Brunswick, Department of Post-secondary Education, Training and Labour Apprenticeship and Certification. (2003). *What is GED?*. Retrieved 1 January, 2007 from http://www.gnb.ca/0381/GED/001ae.htm

Leung, R., Lumsden, J., & Fritz, J. (2006, July 14-16). Accommodating special needs users in the evaluation of an m-learning application: A case study. In *IADIS Mobile Learning Conference*, (pp. 177-184). Dublin, Ireland.

LoPresti, E. F., Mihailidis, A., & Kirsch, N. (2004). Assistive technology for cognitive rehabilitation: State of the art. *Neuropsychological Rehabilitation*, *14*(1/2), 5-39.

LoPresti, E. F., & Willkomm, T. (1997, June). Comparison of commercially available electronic prospective memory aids. In *Rehabilitation Engineering Society of North America Annual Conference (RESNA'1997)* (pp. 523-525). Arlington.

Lumsden, J., Kondratova, I., & Langton, N. (2006, June 29). Bringing a construction site into the lab: A context-relevant lab-based evaluation of a multimodal mobile application. In *1st International Workshop on Multimodal and Pervasive Services (MAPS'2006)*. Lyon, France.

Lumsden, J., Leung, R., & Fritz, J. (2005, June 28-30). Designing a mobile transcriber application for adult literacy education: A case study. In *IADIS International Conference on Mobile Learning*, (pp. 16-23). Qawra, Malta.

Mihailidis, A., & Fernie, G. (2002). Context-aware assistive devices for older adults with dementia. *Gerontechnology*, *2*(2), 173-189.

Moffatt, K., McGrenere, J., Purves, B., & Klawe, M. (2004, April 24-29). The participatory design of a sound and image enhanced daily planner for people with aphasia. In *Conference on Human Factors in Computing Systems (CHI'2004)* (pp. 407-414). Vienna.

Muller, M. (1992, May 3-7). Retrospective on a year of participatory design using the PICTIVE technique. In *Human Factors of Computing Systems (CHI'92)* (pp. 455-462). Monterey, CA.

Myers, B. A., Wobbrock, J. O., Yang, S., Yeung, B., Nichols, J., & Miller, R. (2002, July 15-17). Using handhelds to help people with motor impairments. In *Fifth International ACM SIGCAPH Conference on Assistive Technologies (ASSETS'2002)* (pp. 89-96). Edinburgh, Scotland.

Newell, A. F., & Gregor, P. (2000, November 16-17). User sensitive inclusive design—In search of a new paradigm. In *Conference on Universal Usability*, (pp. 39-44). Arlington, USA.

Stevens, R. D., & Edwards, A. D. N. (1996, April 11-12). An approach to the evaluation of assistive technology. In *ACM Conference on Assistive Technologies (ASSETS'1996)* (pp. 64-71). Vancouver, Canada.

Sullivan, A. F., Jr, & Fischer, G. (2003, June 19-23). Mobile architectures and prototypes to assist persons with cognitive disabilities using public transportation. In *26th Rehabilitation Engineering Society of North America Annual Conference (RESNA'2003)*. Atlanta, Georgia.

Tee, K., Moffatt, K., Findlater, L., MacGregor, E., McGrenere, J., Purves, B., & Fels, S. (2005, April 2–7). A visual recipe book for persons with language impairments. In *Human Factors in Computing Systems (CHI'2005)* (pp. 501-510). Portland, USA.

UNSECO. (2000). *Trouble with reading*. Retrieved 23 January, 2006, from http://portal.unesco.org/education/en/ev.php-URL_ID=8522&URL_DO=DO_TOPIC&URL_SECTION=201.html

Wu, M., Baecker, R., & Richards, B. (2005, April 2–7). Participatory design of an orientation aid for amnesics. In *Human Factors in Computing Systems (CHI'2005)* (pp. 511-520). Portland.

Wu, M., Richards, B., & Baecker, R. (2004, July 27–31). Participatory design with individuals who have amnesia. In *Participatory Design Conference (PDC'04)* (pp. 214-223). Toronto, Canada.

KEY TERMS

Assistive Mobile Technology: Applications running on mobile technology (such as personal digital assistants (PDAs)) that are employed to assist individuals with disabilities in terms of, for example, rehabilitation, and/or to perform daily activities affected by their disability.

Domain Experts: Persons who, as a result of training and/or experience, possess expertise in a field for which technology is being developed.

Evaluation Methods: Techniques used to assess (a) the technological needs of a given user community and (b) the usability and suitability of technological applications.

Guidelines: A set of heuristics, typically based on practical experience, designed to guide a process or development activity.

Individuals with Disabilities: Persons with, for example, physical, cognitive, developmental, psychiatric, or learning disabilities, as well as persons with other special needs, such as the elderly (age 65+) and adults with limited literacy skills.

Participatory Design (PD): A design method in which technology experts, domain experts, and target users work collaboratively to design a solution.

Chapter XXXVII
Mobile Design for Older Adults

Katie A. Siek
University of Colorado at Boulder, USA

ABSTRACT

The global population of older people is steadily growing and challenging researchers in the human computer interaction community to design technologies to help them remain independent and preserve their quality of life. Researchers are addressing this challenge by creating assistive technology solutions using information appliances, such as personal digital assistants and mobile phones. Some have questioned whether older people can use information appliances because of age related problems. This chapter discusses work related to designing, implementing, and evaluating mobile applications for the aging. A discussion about what researchers should consider during the design process for information appliances shows the unique challenges posed by this population.

INTRODUCTION

Our world population is aging. The United States National Institute of Health estimates that the global older adult[1] population grows by 795,000 each month. They project that by 2030, the global older population will grow by 847,000 per month (Kinsella & Velkoff, 2001). In response to this increase, researchers in human computer interaction, social sciences, and ubiquitous computing communities are developing applications to help older people live independent and productive lives. Researchers use *information appliances* (Norman, 1999), such as personal digital assistants (PDAs) (Carmien, DePaula, Gorman, & Kintsch, 2004; Coroama & Rothenbacher, 2003) and mobile phones (Helal, Giraldo, Kaddoura, & Lee, 2003), to create *assistive technologies* for older people.

We contend that older adults can use information appliances if the physical and virtual interfaces are designed to meet their varying needs. Some may argue that older adults do not use information appliances and thus, researchers do not have to adjust designs for this population. However, a recent report in the United Kingdom revealed that 49% of older adults own a mobile phone and of that group, 82% make one or more calls per week (Office of Communications [OfCom], 2006). Thus, older adults are using information appliances, but they do encounter numerous problems, such as font and icon readability and interface complexity issues, discussed in greater detail in the background section.

Other people argue that since younger adults use information appliances now, they will not have a problem using similar technology in the future. Indeed, 82% of all United Kingdom residents own a mobile phone, whereas only 36% of people over 75 years old own a mobile phone (Office of Communications [OfCom], 2006). However, we know that (1) as people age their physical and cognitive abilities do not remain constant and (2) the *digital divide* is still present; factors such as age, socioeconomic status, and disabilities affect individuals' access to technology. Although *walk-up-and-use* systems are becoming more prevalent in our everyday lives, we cannot assume that by giving older people new technology, they will be able to easily interact with the device and application. We must work together now to create a set of guidelines to help inform the design and development of future technologies for older people to avoid problems associated with *technology determinism* (Warschauer, 2003).

In this chapter, we discuss issues that must be addressed when designing information appliance interfaces for older adults. We begin by highlighting design related work with older people and technology - traditional computers and information appliances. We then discuss best practices for conducting user studies with older populations and design issues to consider when developing applications and devices. We conclude the chapter with ideas for future work and challenges to the design, interaction, and technical communities.

BACKGROUND

We discuss how older people interact with traditional computers and information appliances in this section. The related work delves into design and interaction studies because interactions, physical and cognitive, have a major influence on design. Researchers have looked at how older populations interact with traditional desktop computers. Researchers are just beginning to look at how older populations interact with information appliances.

There has been a proliferation of information appliances designed for the general public, including PDAs, mobile phones, remote controls, digital cameras, digital music players, and game playing devices. The interfaces to these vary considerably, suggesting there may be variable age-related performance effects. Hence, when creating applications for older populations, designers must consider age-related abilities such as vision, dexterity, coordination, and cognition. Researchers have discovered that within older populations, there are noticeable differences in abilities, and that different design methodologies, such as universal design (Abascal & Civit, 2001) and user sensitive inclusive design (Newell & Gregor, 2001) should be used. Here we discuss some of the research that has been done to better understand older populations' interaction with technology.

Older People and Traditional Computers

Bernard, Liao, and Mills (2001) found that older people could read faster with a larger, more legible 14-point san serif font on websites. Researchers at Georgia Tech studied how multimodal feedback (sound, touch, visual effect) could assist participants with varying vision problems perform basic mouse tasks (drag and drop). They found that all groups performed better when sound was added; however, groups performed the best when all three modal feedbacks were used (Jacko, Scott, Sainfort, Barnard, Edwards, Emery, et al., 2003).

A number of recent studies focused on the ability of older populations to use PC input devices (Chaparro, Bohan, Fernandez, & Choi, 1999; Charness, Bosman, & Elliott, 1995; Laursen, Jensen, & Ratkevicius, 2001; Smith, Sharit, & Czaja, 1999). The studies showed that older people completed tasks slower than younger groups. Charness et al. (1995) evaluated control key, mouse, and light-pen input devices and found older people preferred the light pen, followed by the mouse and control keys.

Smith et al. (1999) and Laursen et al. (2001) found older people made more mistakes than younger people and had difficulty with fine motor control tasks such as double clicking. Chaparro et al. (1999) found older people performed "point and click" and "click and drag" tasks slower than younger people, but with the same amount of accuracy. The researchers believed the reason that older people were slower was because of reduced fine motor control, muscle strength, and pincher strength associated with older age.

Older People and Information Appliances

Most of the human computer interaction studies on older adults and technology focus on the usability of traditional desktop computers. The usability of information appliances will be scrutinized more carefully as pervasive computing technology applications become more widespread. Researchers are already assessing the needs of older people with respect to mobile phones.

Maguire and Osman (2003) found that older people primarily considered mobile phones as a way to assist in emergencies, whereas younger people saw mobile phones as a way to interact socially. Older people were interested in small phones with large buttons and location aware systems. More specifically, older women were interested in finding the nearest retail shop that met their needs with location aware systems, whereas older men wanted to know how to get places with various forms of transportation. Abascal and Civit (2001) looked at the pros and cons of older adults using mobile phones. They found that older adults liked the safety and increased autonomy mobile phones gave them. But, they were primarily concerned about social isolation and loss of privacy by using a mobile phone. Sri Hastuti Kurniawan (2006) found that older women felt safer with a mobile phone. Unlike younger counterparts, older women wanted brightly colored, bulkier phones with an antenna so it would be easily identifiable in a cluttered purse.

Ziefle and Bay (2005) looked at the cognitive complexity of older adults using mobile phones. They found that older adults performed just as well as younger adults on less cognitively complex mobile phones. They also reported that as the mobile phone interaction became more complex, older participants' performance suffered. Irie, Matsunaga, and Nagano (2005) created a mobile phone for elders by relying heavily on speech input technologies to help decrease complexity and input methods.

Most of the findings in these studies for mobile phones can apply to PDAs as well; however, the needs assessments differ because PDAs have larger physical interfaces and different input mechanisms. The lack of research in the area of PDA technology use by the older adults prompted Darroch, Goodman, Brewster, and Gray (2005) to evaluate a suitable font size for older people who needed to read text on a PDA screen. They found older people preferred reading 12-point font on PDAs, but could read fonts as small as 10 points. The authors pointed out that the lower resolution of their PDA screen could account for the smaller font size preferred by participants than what Bernard and colleagues had previously reported. We looked at how older adults physically interacted with PDAs. We found that older adults had no problem pushing buttons, identifying icons, voice recording, or barcode scanning. Similar to the Darroch study, we found that although older participants preferred to read icons 25-mm large, they could read icons less than 15-mm large (Siek, Rogers, & Connelly, 2005).

Researchers must take into consideration what drives older adults to adopt new technologies for assistive applications to help the target population. Melenhorst, Rogers, and Caylor (2001) found that older adults must understand the benefits of information appliances and alternative communication mediums before they will consider the necessary training to use new technology. In addition, researchers found that for older adults to adopt a new technology, they must feel the technology is useful, convenient, safe, and simple to use, especially in older adults with varying cognitive and physiological abilities (Smither & Braun, 1994).

The findings from this body of research suggest that older people can use information appliances; however, designers and researchers must look at these findings to help inform their designs. More specifically, researchers must look at the physical device capabilities, interface design, and interaction techniques.

MAIN FOCUS OF THE CHAPTER

In this chapter, we broadly define older adults as people over 65 years old. It is difficult to define an ideal older adult because of the variability in older populations' abilities affected by age, illness, and cognitive or physiological decline. Thus, when designing for older populations, it is important to carefully define the target population, recruit older adults who meet the defined criteria, conduct meaningful requirements gathering and user studies, and design *prototypes* with older adults in mind. Here we discuss each of these items in more detail from our experiences in developing assistive applications for older adults.

Recruiting Older Target Populations

The first thing designers and researchers must figure out is what type of older population they would like to target. Will the application or device be for older people with cognitive impairments? Will it be for older people with physical disabilities? Or will the design be for *all* older people? Eisma and colleagues (Eisma, Dickinson, Goodman, Mival, Syme, & Tiwari, 2003) recommend bringing in older people early on in the design process to assist with requirements gathering and prototype development. They found that the different backgrounds of older people and designers mutually inspired the group to create realistic aims for the project. Older people on the design team can help answer questions specific about the abilities of the targeted population. Researchers must keep in mind that if the design is for all older people, the target population will have to be large enough to test people with varying physical, mental, and social abilities.

Researchers typically post fliers, e-mail calls for participations on mailing lists, and recruit participants from their work or university. This may not be the best way to find a pool of older adult participants. Older participants may not have the same *social networks* as the researchers. Thus, researchers should branch out and connect with community centers, religious groups, veteran meetings, assisted living centers, disability support groups, alumni associations, or adult communities to recruit an older diverse population. Typically, researchers can set up a meeting with the activities coordinator, technology group, or outreach liaison to meet older adults.

I would have no need for one of these, so I don't have to touch it. [PDA handed to audience member] But, what if I break it? [Grabs PDA more confidently after researcher says she does not have to worry about breaking it. Pushes a few buttons on the screen.] Well look at that – I could show pictures to my friends.

– Audience member speaking to presenter after recruiting presentation

Similar to any participant population, older adults want to know what is expected of them and what the researcher will do with the data. When recruiting older participants, it is easiest to volunteer to give a presentation about the intended study that includes why the research is being done, what type of person you are looking for (e.g., user profile data), what the participant will have to do, and how the data will be used. The researchers can field questions from the audience to assuage future participants' concerns. Presentations are also the perfect time to hand out preliminary questionnaires to audience members and schedule future meetings for *focus groups, interviews*, or user studies. If participants are expected to use technology that may be unfamiliar to them, bring along the information appliance and let audience members play with the technology after the presentation. Emphasize that you are not testing the participants, but the device or application, and that the device or application cannot be broken with simple interactions. Guided hands-on interactions can change a person's view of the technology as shown in the audience member quoted previously.

Meeting with Older Adults

Designers and researchers will inevitably have to meet with the older adults in their target population during requirements gathering and user studies. There has been quite a bit of research (Eisma et al., 2003; Kurniawan, 2006; Zajicek, 2004) that looks into the best way to meet with older adults. Focus groups and semistructured interview sessions are the most popular meeting methods for requirements gathering and user studies. In this section, we briefly summarize the pros and cons of each method and give tips for best practices.

I have my walking group at 9, craft group at 10:30, doctors at 11:30, lunch at 1…

– Participant and facilitator attempting to schedule another meeting time

A common misconception is that older adults have plenty of time to meet with designers and researchers because they may be retired or work fewer hours. However, researchers may quickly find that some older adults have equally busy schedules. Taking notes about what each person is interested in based on the person's schedule can give insight into how the information appliance would fit into the person's everyday life. The quote about scheduling a meeting shows the participant's varied activities. Would the information appliance always be with her/him during the study? If so, how would she/he carry it when attending each meeting? If not, how can we remind her/him to bring the information appliance to only certain activities? We found that older adults with lower social-economic status have busy schedules too because they were more likely to have chronic illness or responsible for caring for family members.

Participant 1: I do not understand what you are saying. I have to see your lips!

Participant 2: I cannot see the screen because of glare.

– Participants' comments during a focus group

Focus groups typically allow researchers to get peoples' opinions, test ideas for specifications, evaluate prototypes, and learn more from the group by spontaneous discussions. Researchers are divided on how beneficial focus groups are when working with older adults. Zajicek (2004) found that focus groups with over three older people are challenging because of hearing impairments, visual impairments, cognitive abilities, and the ability to follow a conversation. Kurniawan (2006) reported no problems and found that focus groups with over three older people tended to work together and help with *cooperative learning* exercises.

For prototypes that run on information appliances, we found focus groups challenging because of screen glare problems and complex interactions. Information appliances are small; thus, when trying to show a feature or explain an interaction, it is difficult to show it to all participants at once. We have issued each participant an information appliance in focus groups, with multiple researchers on hand to help the facilitator explain concepts, interface components, and interactions. This method allows the participants to see the proposed application and associated small interface components. It also gives the participants the chance to interact with the device and see how input methods are different for information appliances than with traditional desktop computers. Participants typically talked with the facilitator or to the people next to them to compare what they saw and discuss what they thought. Unfortunately, this type of focus group requires more time, preparation, and coaching by researchers. In addition, time must be set-aside for the group to discuss their ideas about the information appliance or application.

Alternatively, we have projected the interface or device onto a larger viewing surface so everyone sees the screen and can discuss the issue at hand. Participants were more likely to talk openly and start new discussions about interface components. The latter method allowed us to guide the discussion more efficiently, but it did not give the participants' the same realistic feel for the interface with smaller buttons and less-controlled interactions as the former method. Designers will have to take into consideration the focus group interaction method to receive appropriate feedback from participants.

Interviews allow the facilitator to work one-on-one with a participant and ask more in-depth questions, or evaluate applications and devices more carefully. We found that we get the most detailed information about interface usability during semistructured interviews with accompanying task-centered user studies. The interview typically is quieter and has fewer distractions for the older user. In addition, the older user has a chance to interact with the device without worrying what others may think of him for not knowing how to do something on the information appliance.

My daughter thinks I am not smart because I cannot use a computer. But you know what—my daughter is not as smart as she thinks she is. One time when she was twelve, she came home from school and…

– Participant comment during interview

Our main problem with interviews, and sometimes with focus groups, is keeping on schedule. Older participants are more likely to share stories with the facilitator about their feelings towards technology when interacting in a one-on-one session. This rich data is useful, but there is a fine balance between keeping the conversation going and making sure the conversation does not diverge too much from the subject at hand, as shown in the previous quote. Another problem we encountered is that older people are more determined to finish each task than their younger counterparts, and will spend extra time to complete the tasks. We found that one-on-one interviews typically lasted one third longer than when working with younger participants.

Physical Interfaces

In this section, we discuss some basic guidelines for the physical design of information appliances based on related work and our experiences. We found that older populations are interested in somewhat larger, more colorful physical devices and input components, although designers must find a balance between size and the perception of size.

When researchers conduct ethnographic studies or conduct studies where technology is discussed, but may not be necessarily used for data analysis, we find a persistent theme; most older adult populations want larger information appliances and input components (e.g., buttons, track wheels, etc.). A larger, bulky information appliance is easier to find, identify, and hold in one hand. Larger input components allow for quicker input. Indeed, a study found that older populations would prefer less overall functionality in exchange for larger buttons (Kurniawan, 2006). In terms of output, older adults would prefer to see a screen with more colors or contrast rather than have a larger screen.

In contrast, when studies have participants interact with the information appliance, they find that bigger is not always better. For example, participants in our studies were worried that their large fingers would press more than one button on an information appliance. The participants soon found that their perception of size was unfounded; they were able to interact with the smaller interface components (Siek et al., 2005). Another study found that older populations with specific physical ailments, such as paralysis, preferred smaller information appliances so they could be tucked into pockets easier (Eisma et al., 2003).

Figure 1. Example of icon sizes older adults can view (preferred size and smallest viewable size) on a PDA

Since older populations are so diverse in abilities, it is difficult to create a strict guideline that specifies criteria of older adults who can use the information appliance. Instead, we have adopted an informal method of bringing information appliances to recruitment meetings and watching how older people interact with the devices. When we give an individual an information appliance, we collect her/his preliminary questionnaire and record comments about how she/he interacted with the device. After the recruitment meeting, the design team meets to discuss the interactions and questionnaire data to make correlations. Occasionally, we invite a clinician or an older adult to help us make conclusions about criteria needed to use the information appliance.

Virtual Interfaces

Similar to physical information appliance design, older adults are interested in the size of interface components and text. In addition, they prefer more common terminology to assist with interactions. Something that has not been studied with information appliances is cognitive interactions and interactions with small widgets and interface components. In this section, we briefly discuss virtual interface guidelines that should be considered when designing information appliance applications for older people.

Older populations typically prefer larger fonts (e.g., 12-point font) (Darroch et al, 2005; Kurniawan, 2006) and icon sizes (e.g., 25mm) (Siek et al., 2005), as shown in Figure 1, but can read much smaller fonts (e.g., 10-point font) and icon sizes (e.g., 15mm). Design teams should take this information into consideration if they prefer to display interface information with text and icons. An application can be more appealing to older adults by using their font and icon size preferences; however, excess scrolling could make the application too complex. Indeed, we have found older populations have difficulty understanding the concept of scrolling on traditional computer Web browsers.

Besides the size of icons, older populations prefer realistic, picture-quality renderings to portray information in icons (Siek et al., 2005). Older participants prefer more detailed icons because the details helped them identify the function of the icon more efficiently.

Audience member 1: Why do I have to press Start to turn off my computer?

Audience member 2: Why do I have to press an apple to turn off my computer?

– Audience members' questions after recruiting presentation

Terminology used in virtual interfaces and user guides are often confusing to the general public. We found older users are more likely to voice their concerns and confusion about terminology. As the previous quote shows, audience members asked simple questions about the Windows and Apple desktop interfaces. At first it stumped the researcher; the reason why we press start and an apple symbol is because we always have. But just because we always have does not mean it is correct. If you would like to turn off your computer and in affect *end* all programs, why would one press start? These questions quickly prompted others to voice their concerns about e-mail and cell phone terminology. Researchers have documented older adults confusion about three-letter acronyms (e.g., SMS, MMS, etc.) and mobile phone terminology (e.g., What is a cell? What is roaming if I am always moving with a mobile phone?).

In this section, we described best practices that helped us and our fellow researchers develop successful information appliance applications for older adults. Since the older adults could use the applications and adopted them in their everyday lives, we assume these practices will help other researchers. We discuss in the next section future research directions for interface input components and interactions with these components. Research in this area will provide practitioners with guidelines to make consistently successful design decisions for information appliances.

FUTURE TRENDS

Information appliances are relatively new technologies, and mobile applications geared strictly towards older adult populations are only beginning to emerge. Researchers will continue to develop assistive applications for elders because of increases in the global older adult population. We must continue to address the issues proposed in this section to help further develop a guideline for information appliance development for older populations.

Researchers know how large common text and icons should be. We must look at how older populations use standard interface component widgets. Can they use standard size widgets with decreased fine motor skills? How large should the widget be?

I can only text people on my mobile phone if they text me first. I just push the reply button. I do not understand how to use the address book or how to enter people's names.

– Participant during interview

Cognitive interactions and interface complexity have been studied with traditional computers and Web sites. Currently, researchers have not delved into these issues for information appliances. Since information appliances have smaller screens and limited input capabilities, there will naturally be more interface screens and with it, increased complexity. As the participant noted in the quote, text messaging on mobile phones requires the user to input data from multiple sources (e.g., address book or alphanumeric key strokes) and send the message. However, once someone has sent a text message, it is easier to push one button and reply to the message. This interaction pattern could be a motivator for future research. How can we use this idea of one button interaction or precached contact data to increase communication mediums for older adults?

Interactions between the physical device and interface components are another area that must be researched for older adults to effectively use information appliances. For example, Charness and colleagues found that older adults had difficulties with traditional computer mouse and directional keyboard input because of varying fine motor control skills and the mapping between lateral movements with the mouse and the coordinate system on the screen. They found that light pens were optimal for older adults (Charness et al., 1995). In terms of information appliances, PDA screen input is similar to a light pen for optimal input. Despite this connection, designers must take into consideration that older adults may not have the fine motor control needed to select the standard, tiny interface components on PDA screens. In addition, current mobile phones pose an even bigger challenge, given the directional key presses needed to scroll and input information. It would be interesting to study if having these interactions close to the screen and on the same coordinate plane, such as in information appliances, will affect older peoples' perception of ease with information appliance input.

Along with standard interface development, researchers must strive to diversify the pool of older participants in their studies. Most studies summarized in this chapter worked with educated older populations. Indeed, recruitment from private assisted-living communities is fairly easy because the older adults who live in the community are educated and curious about technology. But a pessimistic view of the future may be that with such a large, ever growing population of older adults, the people who cannot afford private care will be monitored remotely by information appliances and *context aware* systems. If we create design guidelines and information appliance systems tested by people who are comfortable with technology, then we are leaving out the population who may need to use this technology one day. Researchers and designers must try to diversify their user pools by looking at education and socioeconomic status of their participants.

CONCLUSION

In this chapter, we looked at current research conducted with older adult populations using traditional desktop computers and information appliances. Research in the area of interface design for older adults is deficient because information appliances are relatively new, and design of assistive applications for older adults is just beginning to mature. We discussed issues and best practices that must be addressed when designing for information appliances. More specifically, we looked at the diversity of older adults, recruiting target populations, meeting with older adults in focus groups and interviews, and physical and virtual interface design considerations. We feel these best practices are useful for researchers and the general practitioner because of our success with developing applications for older adults. Researchers and designers must strive to diversify their older adult target populations and consider people with different physical, cognitive, and emotional abilities. In addition, people from varying socioeconomic groups must be considered for the study to see how computing experience affects performance with information appliances.

REFERENCES

Abascal, J., & Civit, A. (2001). Universal access to mobile telephony as a way to enhance the autonomy of elderly people. In *Proceedings of the 2001 EC/NSF Workshop on Universal Accessibility of Ubiquitous Computing: Providing for the Elderly* (pp. 93-99). New York, NY: ACM Press.

Bernard, M., Liao, C. H., & Mills, M. (2001). The effects of font type and size on the legibility and reading time of online text by older adults. In *CHI '01 Extended Abstracts on Human Factors in Computing Systems* (pp. 175-176). New York, NY: ACM Press.

Carmien, S., DePaula, R., Gorman, A., & Kintsch, A. (2004). Increasing workplace independence for people with cognitive disabilities by leveraging distributed cognition among caregivers and clients. *Computer Supported Cooperative Work*, *13*(5-6), 443-470.

Chaparro, A., Bohan, M., Fernandez, J., & Choi, S. (1999). The impact of age on computer input device - Psychophysical and psychological measures. *International Journal of Industrial Ergonomics*, *24*(5), 503-513.

Charness, N., Bosman, E. A., & Elliott, R. G. (1995). *Senior-friendly input devices: Is the pen mightier than the mouse?* Paper presented at the 103rd Annual Convention of the American Psychological Association Meeting, New York.

Coroama, V., & Rothenbacher, F. (2003). The chatty environment - Providing everyday independence to the visually impaired. In *UbiHealth 2003.*

Darroch, I., Goodman, J., Brewster, S. A., & Gray, P. D. (2005). The effect of age and font size on reading text on handheld computers. In *Lecture Notes in Computer Science: Human-Computer Interaction - Interact 2005* (pp. 253-266). Berlin/Heidelberg, Germany: Springer.

Eisma, R., Dickinson, A., Goodman, J., Mival, O., Syme, A., & Tiwari, L. (2003). Mutual inspiration in the development of new technology for older people. In *Proceedings of INCLUDE 2003* (pp. 7:252-7:259). London, United Kingdom.

Helal, S., Giraldo, C., Kaddoura, Y., & Lee, C. (2003, October). Smart phone based cognitive assistant. In *UbiHealth 2003*.

Irie, T., Matsunaga, K., & Nagano, Y. (2005). University design activities for mobile phone: Raku Raku PHONE. *Fujitsu Sci. Tech. J., 41*(1), 78-85.

Jacko, J. A., Scott, I. U., Sainfort, F., Barnard, L., Edwards, P. J., Emery, V. K., et al. (2003). Older adults and visual impairment: What do exposure times and accuracy tell us about performance gains associated with multimodal feedback? In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 33-40). New York, NY: ACM Press.

Kinsella, K., & Velkoff, V. A. (2001). *An aging world: 2001*(U.S. Census Bureau, Series P95/01-1). Washington, DC: U.S. Government Printing Office.

Kurniawan, S. (2006). An exploratory study of how older women use mobile phones. In *Proceedings of UbiComp 2006: Ubiquitous Computing* (pp. 105-122). New York, NY: ACM Press.

Laursen, B., Jensen, B. R., & Ratkevicius, A. (2001). Performance and muscle activity during computer mouse tasks in young and elderly adults. *European Journal of Applied Physiology, 25*, 167-183.

Maguire, M., & Osman, Z. (2003). Designing for older inexperienced mobile phone users. In *Proceedings of HCI International 2003* (pp. 22-27), Mahwah, New Jersey: Lawrence Erlbaum Associates.

Melenhorst, A.-S., Rogers, W. A., & Caylor, E. C. (2001). The use of communication technologies by older adults: Exploring the benefits from the user's perspective. In *Proceedings of the Human Factors and Ergonomics Society 45th Annual Meeting* (pp. 221-225).

Newell, A. F., & Gregor, P. (2001). Accessibility and interfaces for older people - A unique, but many faceted problem. In EC/NSF *Workshop on Universal Accessibility of Ubiquitous Computing: Providing for the Elderly.*

Norman, D. (1999). *The invisible computer: Why good products can fail, the personal computer is so complex, and information appliances are the solution*. Boston: MIT Press.

Office of Communications. (2006). *Media literacy audit: Report on media literacy amongst older people*. London, United Kingdom: OfCom.

Siek, K. A., Rogers, Y., & Connelly, K. H. (2005). Fat finger worries: How older and younger users physically interact with PDAs. In *Lecture Notes in Computer Science: Human-Computer Interaction - Interact 2005* (pp. 267-280). Berlin/Heidelberg, Germany: Springer.

Smith, M. W., Sharit, J., & Czaja, S. J. (1999). Age, motor control, and the performance of computer mouse tasks. *Human Factors, 41*(3), 389-396.

Smither, J. A., & Braun, C. C. (1994). Technology and older adults: Factors affecting adoption of automatic teller machines. *The Journal of General Psychology, 121*(4), 381-389.

Warschauer, M. (2003). Demystifying the digital divide. *Scientific American*, 42-47.

Zajicek, M. (2004). Successful and available: Interface design exemplars for older users. *Interacting with Computers, 16*, 411-430.

Ziefle, M., & Bay, S. (2005). How older adults meet complexity: Aging effects on the usability of different mobile users. *Behaviour and Information Technology, 24*(5), 375-389.

KEY TERMS

Assistive Technologies: Applications and devices that pair human computer interaction techniques and technology to enhance the quality of life for people with various special needs.

Context Aware Systems: Technology embedded into our environments that communicates location, action, and other variables to help monitor the environment or individual.

Cooperative Learning: A method that allows individuals with different abilities to work together to improve their understanding of a subject.

Digital Divide: The gap between groups of people who do and do not have access to information technology.

Information Appliances: Electronic devices that allow people to send and receive various types of media (e.g., PDAs, mobile phones).

Focus Groups: A small group of selected participants who are asked questions about what they think about a specific topic or product (e.g., prototype); participants are free to discuss and build on what other participants say.

Interview: A participant is asked a series of questions by a facilitator to learn the participant's personal thoughts about a topic or product (e.g., prototype). Facilitators ask more open-ended questions in semistructured interviews and adapt future questions based on participants' feedback.

Prototype: A software or paper-based system that has a subset of the final application functionality; integral part of software development that allows researchers to get feedback from users before developing a fully functional system

Social Network: Connections between individuals with personal and professional relationships. Often the strength of the connections and influences of relationships are taken into account.

Technology Determinism: Idea that by introducing technology, people will understand and be able to use it.

Walk-Up-and-Use: Technologies that allow people to use the device or application without previous training or instruction (e.g., bank machines, self check-out kiosks at stores).

ENDNOTE

1 Older people are defined here as 65 years old and over.

Chapter XXXVIII
Designing Mobile Applications to Support Mental Health Interventions

Mark Matthews
Trinity College Dublin, Ireland

Gavin Doherty
Trinity College Dublin, Ireland

David Coyle
Trinity College Dublin, Ireland

John Sharry
Department of Child and Family Psychiatry, Mater Hospital, Ireland

ABSTRACT

The advent of mobile technology has brought computing to a wide range of new contexts, some of which are highly sensitive and place new constraints on the designer. In this chapter we discuss issues related to the design and evaluation of mobile software for sensitive situations, where access to the end user is extremely restricted. We focus on the specific example of technological interventions that support adolescents in mental health care settings. We examine the practical and ethical constraints placed on access to end users and contexts of use, and how this may affect approaches to design and evaluation. General design recommendations for this area are described. We consider approaches to iterative design with mental health care professionals, and how research on technological and therapeutic aspects may proceed in tandem. We identify methods that can be used when conducting evaluation in these limited situations and describe a methodology for maximising the value of such evaluation. By way of illustration, we present the design and evaluation of a mobile phone-based "mood diary" application designed for use in clinical situations by adolescents undergoing mental health interventions.

INTRODUCTION

There are a variety of well-established methods for the design and evaluation of user interfaces. Recent developments in mobile technology have brought technology into a variety of novel contexts. Because of this, there has been a focus on ethnographic and context-based approaches to requirements gathering and design. While there are new challenges posed by these contextual factors, by and large, designers can draw on adapted versions of tried and tested methods for requirements elicitation, design, usability improvement, and evaluation. Hence, questionnaires and user interviews, paper prototyping, user observation, think aloud protocols, and so on, can be applied in the design and development of new mobile systems.

In some situations, established methods are not viable because of the sensitivity of the environment or the end users. Sensitive situations, as discussed in this chapter, are situations where poor quality solutions may have a highly negative impact, and where the introduction of not just the technology, but also the designer, could be detrimental to the environment that is the target of the technological intervention. As a result of this sensitivity, one could have a system under development where there is no direct access to end users. For such situations, it is necessary to identify new methods to address the problems of design and evaluation in these circumstances. This chapter's main focus is on identifying methods for design and evaluation in sensitive situations where access to the end user is extremely restricted. To achieve this, the chapter concentrates on one particular situation; the use of mobile software in talk-based mental health care with adolescents. An ancillary aim is to give sufficient background detail to help readers understand the significant constraints present in this area. Many of the issues raised are applicable to a wider range of sensitive situations.

This background section begins by briefly describing why adolescent mental health is a particularly sensitive situation, and discussing why mobile software has particular promise in this environment. It continues with a description of the ethical constraints the designer needs to consider. We then discuss which traditional usability evaluation methods are applicable in this situation, and discuss related research in this and similar domains. Finally, we present a characterization of the end users; both the adolescent clients and the therapists. The chapter then moves on to present a range of design and evaluation recommendations. In particular, we present a multistage methodology for developing mobile software in this sensitive situation, and end by describing a case study that demonstrates the benefits of this process.

Mental Health: A Sensitive Situation

Mental health is one of the most pressing concerns for public healthcare systems worldwide. A large-scale international study has identified mental illnesses as the second leading cause of disability and premature mortality in the developed world (Murray & Lopez, 1996). The past 50 years have witnessed major advances in the treatment of mental illnesses. One significant report concluded that the efficacy of mental health treatments is well documented, and that a range of effective treatments exist for most mental disorders (Surgeon General, 1999). Treatment generally takes the form of talking, listening and learning, physical treatment (drugs, ECT, biomedical), and/or social interventions. This chapter focuses on listening-andtalking-based intervention methods. Many studies have concluded that talk-based methods are equally as effective as drug-based interventions in the treatment of many disorders (Surgeon General, 1999).

When designing technology for mental health care (MHC) environments, researchers have to carefully consider and characterize both the adolescent client and the therapist. Both are sensitive to the introduction of new technologies in a therapeutic setting. Adolescence is the transitional period between childhood and adulthood, and is a time when professional support for dealing with mental illness is in greatest need. However, most adolescents with mental health problems do not receive professional help and those who do can find it difficult to engage with the available services (Offer, Howard, Schonert, & Ostrov, 1991). Therapy is typically something imposed upon them and because of this, they are often unwilling to accept it. Difficulties in engaging adolescents can also arise due to the serious nature that therapy is seen to have in society and the stigma that is often attached to it (BMA, 2006). Exploring new ways to engage adolescents in professional therapeutic services is one of the most important and difficult challenges in mental health. By engagement, we mean helping adolescents develop an active interest and confidence in the outcomes of therapy.

Sensitivity is required when dealing with therapists who may view technological solutions with suspicion. In medical healthcare, professionals are often uncomfortable to find themselves in the role of computer novice (Gosbee & Ritchie, 1997). For therapists, this may be the case to an even greater degree, as technology is rarely used in day-to-day work or in therapist training. As a result, they may feel the introduction of technology might undermine their authority. Any technology that is introduced needs to compliment a therapist's current practice and methods, and not demand too much time in an already busy schedule. Small innovations are required to develop confidence in the possible benefits technology can offer.

Why Mobile Phones?

Client self-efficacy and the degree to which clients engage with their treatment has been identified as one of the most significant factors in achieving positive therapeutic outcomes (Assay & Lambert, 1999). Alongside engagement during clinical sessions, a key aim in improving client self-efficacy is to increase the degree to which clients engage in prescribed therapeutic activities between therapeutic sessions. Within public health care systems it is often the case that clients will see therapists, at most, once a week, and sometimes will only attend once a month over the course of treatment. For this reason, it is particularly important to engage clients with the therapeutic process between sessions. Existing (e.g., paper-based) approaches have achieved limited success in improving client engagement between therapeutic sessions. Identified problems with materials currently used include lack of client engagement, stigma associated with therapeutic activities, and inconvenience.

The mobile phone is a promising means for achieving greater client engagement between therapeutic sessions. The mobile phone is the most personal communication device with widespread ownership, particularly among adolescents. A recent UK survey indicated that 90% of 12 year olds owned a mobile phone and sent three times as many text messages as their parents (BBC, 2006). It is an interactive and adolescent-centred technology. Its strengths include ubiquitous and continuous access, familiarity, and ease-of-use. It has the potential to make therapeutic activities more convenient, less stigmatic, and more engaging.

Adolescents are already familiar with mobile phones and how they function: they use them daily to communicate with peers, play games, and send textual and multimedia content. A high proportion of adolescents are familiar with textual input, using it every time they send a text message. They may find entering text on their phones significantly more appealing and comfortable than using pen and paper, which may have negative connotations. This is very important for clients with literacy difficulties. The phone can simplify the transfer of any client content to a secure server, alleviating the client of the responsibility to do so.

SUPPORTING MENTAL HEALTH INTERVENTIONS

To provide context for the case study and shed light on the issues involved, we briefly introduce the most important issues to consider when introducing technology in this area. For an extensive review on the use of computers in mental health interventions see Coyle, Doherty, Matthews, and Sharry (2007).

Most adults are comfortable with face-to-face dialogue. This is often not the case with adolescents. Children struggle to express themselves with words alone, and much research has been conducted into ways of engaging children using indirect channels and play. Adolescents can be resistant to these methods; they like to be treated as adults and will not engage if they perceive they are being treated as a child. Equally, many teenagers are private and self-conscious and often react with confrontation or not at all to direct dialogue with a therapist. Given the general enthusiasm of many adolescents for technology, an opportunity exists to use it to assist with adolescent interventions. Technology may provide a channel through which indirect communication between the therapist and client can proceed in a less confrontational fashion. For example, one study on using a computer game in a therapeutic session, reported that clients found it easier to talk while looking at the screen. (Coyle, Matthews, Sharry, Nisbet, & Doherty, 2005). It also found that the introduction of the game had an empowering effect on the adolescent clients. They were more comfortable with the technology than many of the therapists.

Previous research outlines the central importance of client factors (client and environment strengths, client resources) and the quality of therapeutic alliance (i.e., the relationship between therapist and client) in affecting positive change through psychotherapy (Assay & Lambert, 1999). Therapeutic outcomes are more likely to succeed if the therapist engages in a client-centred manner. An effective therapeutic process should actively engage adolescent participation and involve their interests, strengths, and ideas. Four core values have been identified as fundamental to establishing a beneficial working relationship: respect, empathy, genuineness, and client empowerment (Egan, 2002).

The aims of technological interventions within this area can include acting as an icebreaker, providing material to talk about during the session, reenforcing coping strategies introduced during therapy, reducing the potentially confrontational and stressful dynamic of a client facing a therapist, and providing peer content so that the client does not feel that they are alone in experiencing difficulties. It is worth noting that a technology intervention need not necessarily have a long-term effect. Even an intervention that achieves greater engagement through a "novelty" effect may contribute to the therapeutic process by building a therapeutic relationship or acting as an ice-breaker.

Design Under Ethical and User Access Constraints

Research into the use of technology in talk-based interventions for adolescents must adhere to the strict ethical guidelines of the domain. The sensitivity of this situation is further enhanced by the social stigma often associated with mental illness. Many activities are sensitive and even the most unobtrusive methods of observation by a third party could affect trust and the efficacy of the intervention. All therapists are required to obey the Hippocratic Oath, often summarised by the simple principle "first do no harm." This means that above all, every possible precaution must be taken to ensure that interventions do not have harmful effects on the client. Research into the use of technology must adhere to this strict ethical requirement. Roberts and Dyer have produced a guide to MHC ethics and provide solutions to ethical problems (Roberts & Dyer, 2004). HCI researchers should be aware of these constraints, and with the protocols for evaluation.

Ethical requirements place strict limitations on access to MHC settings. It is unlikely that many HCI researchers will have the required qualifications that would allow direct access to this sensitive setting or for contact with people suffering mental health disorders. It is therefore necessary to find solutions for design and evaluation that take into account these access constraints. Access to therapists is not restricted and they can act as proxies in clinical evaluation. An important element of any approach is the ability to maximize the evaluation that could be conducted independently of clinical settings.

There are practical implications to these constraints. To evaluate software in a mental health environment, a detailed proposal has to be submitted to an ethics board. This typically takes at least a month and possibly longer, should the board require changes to any part of the study. Furthermore, if any substantial changes are made to the software then the ethics review procedure must be performed again. This has a significant impact on the duration of evaluations and reduces the practicality of iterative design with end users. All software must be thoroughly evaluated, before it can be made generally available. Ethical restrictions rule out user-assisted design and participatory design with adolescent end users, and also mean less experimentation is possible.

Designing for Adolescents

The initial problem facing researchers when designing for adolescents in MHC settings is how to characterize them. Since access is extremely limited, we must rely on discussions with therapists and analysis of the research literature.

Adolescents attending therapists suffer from a range of problems including emotional disorders (e.g., anxieties, depression, and phobias), eating disorders, conduct disorders, and self-harm. They are a heterogeneous group and have varying backgrounds, motivations, abilities, and interests. Socioeconomic factors play a major part in the prevalence of mental health problems. A recent survey found that young people living in rented accommodation are twice as likely to suffer from a mental health problem (Office for National Statistics, 2005). Poor opportunities in education and employment are also considered risk factors. Adolescents with learning disabilities are more at risk of developing mental health problems (Royal College of Psychiatrists, 2004). There is a strong sense of stigma surrounding mental illness. Research has identified it as a major concern of adolescents attending therapy (Office of the Deputy Prime Minister, 2004). A particular difficulty is that young people tend to have more discriminatory attitudes to mental health problems than adults.

A recent study has reported that in MHC settings, adolescents need to feel respected, listened to, and not judged by health professionals (BMA, 2006). They strongly value confidentiality, and consistency of care is important. It was found to be essential that the clients do not have to repeat their story to different therapists. It is clear that in respect to these issues, technology can play a valuable role.

Designing for Therapists

Many therapists have pressurized work schedules, and it is important that technologies do not add to this pressure by placing extra demands on their time. An informal survey undertaken by the authors revealed that while many therapists are familiar with Microsoft Office, e-mail and Internet software, few are likely to have had specific training or experience in the use of technology in client contact situations. Although computer-based training has many documented benefits and has become increasingly popular in other health science and educational programs, it is still largely unused in therapist training. These factors have a knock-on effect on the use of computers in clinical practice. When questioned about increased use of technology, many therapists express concerns over the need for additional training. They are concerned that existing skills may become obsolete (Caspar, 2004). At this early stage in the development of technology for therapeutic interventions, it is desirable that systems should be designed to take advantage of the existing skills of therapists and integrate with current working methods.

Many therapists cite fears over the security of sensitive information as one of their primary reasons for scepticism of the benefits of technology (pers. comm.). While many methods have evolved for ensuring the actual security of electronic data, perhaps more significant is the perceived security of this information by both therapists and their clients. In peer support and collaborative group systems it is also essential that the accuracy of information and the trust of individuals be maintained and represented. Maximizing this sense of security and trust is thus an important topic for further HCI research in this area.

It is important that the introduction of technology involves protocols for implementation and use. Such protocols can prevent misunderstandings and ease concerns related to the use of the technology. A particular concern for therapists is that the introduction of a new technology brings greater responsibility and more opportunities for client-therapist contact. For example, some therapists were concerned that the introduction of mobile technology may add to their workload and responsibility by introducing a constant line of communication between them and a client that they would feel obliged to monitor. Designers have to take such concerns into consideration.

Mobile Software to Support Psychotherapy

Adolescents spend less than 1 hour a week on average in a therapeutic session. Mobile technology holds promise for engaging adolescents when they are away from the session. It is a portal into adolescents' personal space, and can offer them persistent and continual access to therapeutic materials and activities outside of the therapy session. It affords the user a level of privacy and security unattainable with other technologies. It is a simple matter to provide a mechanism to secure personal and sensitive information. Using a mobile phone in public is a common and accepted activity. The privacy and security provided by the mobile phone may contribute to greater candour. Research has reported that SMS users are more likely to be intimate and revealing in text messages than in face-to-face conversations (Schaefer & Reid, 2001).

As a networked and an interactive device, the mobile phone can initiate therapeutic activities. Clients can be sent reminders via their phones to perform a particular task. Mobile content can be automatically uploaded to a secure location, which greatly reduces the possibility of the loss of information and subsequent loss of trust in the software. As many adolescents carry their phone almost everywhere they go and do not have to make an extra effort to remember their therapeutic materials, the use of the mobile phone may help to overcome some of the barriers to treatment by providing access to therapeutic materials and activities regardless of location and clinic opening hours. Introducing a device that has many positive connotations for adolescents may also help to offset some of the stigma associated with attending therapeutic sessions. Another advantage in this context is that it gives access to periods where the client is unoccupied or bored, when they might willingly engage in therapeutic activities.

A successful implementation using mobile technology may support the therapeutic process by (1) creating content for use in a session, (2) engaging adolescents in beneficial therapeutic activities, (3) encouraging them to spend more time thinking about therapeutic matters, and (4) making therapeutic material more convenient and accessible.

There has been considerable research into using mobile technologies in education (Naismith, Lonsdale, Vavoula, & Sharples, 2004; Roschelle, Sharples, & Chan, 2005; Savill-Smith, 2005). Some of this research can be relevant to mental health care, as it deals with children in similar technological environments where pen and paper activities are still dominant. It is an area that requires some sensitivity, albeit significantly less than clinical situations. One project used mobile phones with schoolchildren as a bridge to a Web portal (Vavoula, Sharples, Rudman, Lonsdale, & Meek, 2006). The mobile phone was used for data collection on the move around an art gallery. The content collected was later viewed and organized through a Web portal, which allowed for more complicated interaction. This project found that the mobile phone was less fragile and more engaging than the paper alternatives. This project used a process where evaluation and development took place at the same time (Vavoula, Meek, Sharples, Lonsdale, & Rudman, 2006). Outcomes from one stage of the development informed the next stage.

Mobile software has been demonstrated to hold much promise in general health care, potentially supporting more attentive care at a reduced cost and with more comfort to the patient. Text messaging (SMS) has been used to generate daily medication reminders (Neville, Greene, McLeod, Tracy, & Surie, 2002). Systems for monitoring dialysis (Nakamoto, Kawamoto, Tanabe, Nakagawa, Nishida, Akiba, & Suzuki 2003) and asthma (Sanderson, Andrews, Corry, & Lapsley, 2003) have been developed. One project developed a rapid application development methodology for producing personalized educational software for breast cancer patients (Wood, Keen, Basu, & Robertshaw, 2003). However, there is a significant difference between producing general healthcare applications and producing software for users with psychological problems, who may be considered fragile. There is the added complication of stigma and the potential embarrassment associated with mental illness. This is magnified when end users are adolescents. Nevertheless, examining approaches and methodologies from similar areas can have benefits for research in this area.

Research into using the mobile phone as a therapeutic support device is at an initial stage. Few, if any, studies have looked at HCI issues, but have focussed generally on therapeutic evaluation. One study using SMS messages in the aftercare of bulimia nervosa patients outlined the following benefits of mobile technology (Bauer, Percevic, Okon, Meerman, & Kordy, 2003): (1) access is independent of space and time, (2) there is little temporal or monetary cost, (3) it is an interactive medium, (4) patients are encouraged to express themselves briefly and concisely, (5) it allows patients to stay in contact with therapy, and (6) it can help to bridge the gap between inpatient and outpatient treatment.

Using the mobile phone may encourage adolescents to consider therapeutic matters. One study, which used PDAs as an extension to therapy, reported that signals from the device reminded clients to think about things like eating behaviour and mood (Norton, Wonderlich, Myers, Mitchell, & Crosby, 2003).

The mobile phone provides an increasingly powerful and ubiquitous platform with the potential to support a variety of applications. The types of therapeutic activity that are most suited to mobile phones have yet to be established. Functions that are familiar to adolescents (e.g., entering text, taking photos) may have powerful applications in this area, including content collection, storytelling and psychoeducation. There is an opportunity to build on developments with mobile phones in less restricted areas such as education, while considering the particular characteristics of the mental health context, which brings with it a different set of design concerns.

RECOMMENDATIONS

The following sections look at the design and evaluation of mobile software for adolescents and therapists involved in professional mental health services. Traditional methods for design and evaluation must be adapted. We present a number of recommendations based on our experience working in this area. Careful design is essential to ensure that technology acts as an effective support to the therapeutic process and not a barrier.

Interaction Design Methodologies

Existing approaches to interaction design and usability improvement involve a number of stages. These range from high-level requirements gathering and analysis, to low-level, detailed design. A range of techniques is used to improve usability and ensure compatibility with user needs. These techniques involve end users, either directly (as with participatory design), or as a source of information for analysis (as with ethnography and contextual analysis), or to validate and explore designs (often low-fidelity and "rapid" prototypes). Even with analytic approaches (such as task and activity analysis), end-user validation of models is seen as an important aspect.

In recent years, there has been an emphasis on studying applications in the context where they will be used (Suchman, 1987). Researchers and designers looking at interaction in a mobile context may need to employ a combination of new and established data gathering and evaluation methods.

Only relatively recently have children (including adolescents) been studied as a distinct group in HCI (Brukman & Bandlow, 2002). Druin presented a framework describing the roles that children can play in the evaluation of technology, and urges active participation from children as design partners (Druin, 2002). Markopoulus and Bekker have written on the assessment of evaluation methods for children (2003). There are detailed guidelines for HCI researchers working with children (Barendregt & Bekker, 2005; Hanna, Risden, & Alexander, 1997; Hanna, Risden, Czerwinski, & Alexander, 1999). These approaches assume some degree of contact with end users and hence, in this domain, are subject to the same access difficulties as standard HCI techniques.

Given the lack of direct access to end users and the need to test by proxy through therapists, it is unavoidable that the time taken to feed results back from end users is greatly increased. Coupled with the constraints on information gathering, this effectively means that we cannot recommend applying standard approaches to user-centred design without modification. As there is an imbalance in access to the two user groups, the obvious option is to rely heavily on therapist evaluation and opinions on designs. While it is appropriate to see the goal of the designer in terms of providing technology support to therapists, full exploitation of the possibilities of technology within such areas can only come from understanding the relationship and interaction between both types of end users and technology.

Data Collection

In the limited settings of mental health interventions, it is important to gather as much information from evaluations as possible and make the most of the available evaluation methods. Furthermore, because new technology must be evaluated from both a therapeutic and a HCI perspective, there is the risk of overburdening the therapist and client.

It is extremely difficult to obtain fine-grained contextual data on the use of mobile applications by adolescent clients. When software is evaluated in a therapeutic session, the therapist is a witness to how the technology is used by the client. With mobile software, it is difficult to know what context it is being used in. Implementing application logging within mobile applications goes some way to providing interaction information. Available measures include time-stamping interactions, analyzing time spent on task and examining which functions of the software were used and with what frequency. While this process is potentially invasive to the privacy of the adolescent client, the benefits to be gained justify the measure, provided that there are strict protocols for dealing with this information. This information can provide insights into what areas of the software adolescents are spending most time on and any difficulties they may be experiencing. Depending on the purpose of the software, further measures may be possible. For example, if it is primarily designed to gather client content, then appraising the quality and frequency of this content can go some way to evaluating its effectiveness.

In clinical trials, therapists participating in evaluations must agree to act as proxies. They are then responsible for administering pre- and postquestionnaires to their clients. As time is limited, it is not reasonable to expect clients to spend a long time in a session filling out questionnaires. For this reason they have to be succinct and clear. They should allow clients the opportunity to openly express their opinion on the software, as well as gathering information about use of the software. Protocols for data collection should provide clients with privacy and anonymity when completing questionnaires. Therapists must give clients space to express their opinion. In order to allay client concerns that their therapist might read their questionnaires, it is advisable to seal them in envelopes once completed. Due to the constraints of the setting, it is generally not possible to have therapists administer interviews to their clients.

The HCI researcher can ask a little more of participating therapists, who are often enthusiastic, and keen to add to their repertoire of therapeutic tools. Questionnaires and interviews with therapists before and after testing are permissible. The focus is on how the technology affected their usual practice and their impression of how the adolescent perceived the technological intervention. In many cases, therapists may have a client attend over a period of months before, during, and after a technological intervention. This gives them a good perspective to assess the affect the technology had on the client and the intervention.

The evaluation of mobile software for clinical settings is a long process that begins with sending detailed proposals to an ethics board, and involves long delays while therapists wait for the appropriate time to introduce the mobile application into an intervention. Testing can overrun, due to clients' poor attendance and other factors. These delays mean that clinical evaluations cannot reasonably form the basis for an iterative design process.

Information Gathering Methods

The advent of mobile computing has had significant impact on the HCI field, underlining the importance of context. Significant challenges have arisen regarding the conduct of usability studies for mobile systems, particularly concerning the gathering and analysis of data, and the difficulty of isolating contextual factors. Innovative solutions are sometimes required; one non-mental health study examined the use of mobile phones by 16-22 year olds (Carroll, Howard, Peck, & Murphy, 2002). It made use of several evaluation methods including scrapbooks with disposable cameras, online diaries, and focus groups. However, in an MHC setting, the restrictions on access make such rich information gathering difficult; data collection in clinical settings is severely limited and unsatisfactory for the purposes of usability improvement. This affects both requirements analysis and usability improvement activities.

Below is a consideration of some common information gathering methods used in requirements analysis and usability improvement, and the possible use of these methods in sensitive settings such as adolescent mental health care.

- **Interviews and discussion groups:** Discussion groups have been used in non-mental health studies with adolescents to good effect. (See, for example, Grinter and Eldridge 2001). However, due to the ethical constraints, interviews and discussion groups between HCI researchers and adolescents in therapy are effectively impossible. By contrast, interviews with therapists are a valuable source of information.
- **Think-aloud evaluation:** Think-alouds have been used in non-mental health studies with adolescents to examine their interaction with cd-rom encyclopaedias (Branch, 2001). However, they can be problematic with mobile systems, although not impossible with appropriate equipment. For sensitive situations, however, they are subject to the same ethical constraints that curtail the use of video and audio logging.
- **User notebooks:** User notebooks are often used for longitudinal studies, and have the advantage of being applicable to mobile systems, particularly where interaction with a system is relatively sparse and spread over a long period. There are difficulties recording entries in notebooks when the users are on the move (Grinter & Eldridge, 2001). For many sensitive situations, this approach is unlikely to work because of privacy concerns (in some cases related to stigma), or because the end users are unlikely to find completing the notebook entries to be an engaging activity (as with adolescents). One non-mental health HCI study, taking this into consideration, implemented a voice-mail diary technique for participants in a mobile technology study (Leysia & Marilyn, 2002). This may be worth exploring, although it is likely that any HCI evaluation technique that requires adolescent clients to focus more on usability evaluation than therapeutic engagement will not be acceptable. As it stands, it is extremely difficult to engage adolescents in therapeutic activities.
- **Questionnaires:** These can be administered by proxy without loss of confidence in the accuracy of the data. Because of the controlled nature of use, the completion rates can be very high.
- **Application logging:** Software is available to record fine-grained data of users' interactions; for example, cursor movement on desktop computers. The mobile phone, with its limited power and storage capacity, does not support intensive operations such as screen capture. More modest logging is

possible, such as navigation choices and time spent on task. This can be a useful means of gathering data once client and therapist approval has been established. Appropriate permissions and awareness of this data collection must be part of the protocol.

- **User observation:** User observation, in context, often involves either following users and observing their mobile use from afar, or end users wearing special video recording equipment. Both methods are inappropriate in adolescent mental health settings. It may be argued that observation of therapy sessions using video recordings could be used as an alternative to direct contact. However, given the ethical constraints and privacy demanded of therapeutic interventions, even limited second-hand observations are likely to prove ethically sensitive, time consuming to organise, and prohibitive to regular use. Even in the training of therapists there is a reliance on role-play rather than direct contact. While video of real sessions is used in training of therapists, it is not straightforward to receive clearance to make such recordings. Moreover, it is not generally possible for mobile users. A further difference is that video recordings made for training purposes can be reused many times over a period of years. This justifies the effort to overcome the ethical issues involved and is a consideration that does not apply to usability improvement.

Mobile Design Recommendations

The following guidelines are intended as high-level recommendations for the design of mobile software in mental health situations with adolescents:

1. **Ethical requirements:** Research into the use of mobile technology in MHC settings must adhere to strict ethical requirements. To help in meeting these requirements, we suggest that new systems be (1) based on accepted theoretical models of MHC, (2) designed in full collaboration with therapists, (3) designed to integrate with existing working methods, and (4) used by clients under the guidance of a professional therapist. Prior to being used in clinical settings, it is essential that (1) the reliability and usability of any new system is thoroughly verified, (2) the therapeutic validity of systems is evaluated by therapists, and (3) ethical clearance must be agreed for all proposed studies. The process of validating systems and gaining ethical clearance is a time consuming process that must be completed prior to each stage of the clinical evaluation of a system.

   Protocols should be provided for adolescents, therapists, and parents/guardians regarding use of the software. It is particularly important to provide reassurance to the adolescent that they have full control of access to their personal information.
2. **Design for privacy and engagement:** Privacy and security are deeply intertwined for end users. It is a common sight to see people interacting with their phone in public, and so it is straightforward for an adolescent to avoid attracting attention when using mobile software. The small display screen and one-handed text entry provide increased privacy. Mobile software should be easily locked with a pin so the adolescent alone has access to their information, providing a sense of security. Should the adolescent's phone be lost or even taken off them by a peer, there is no fear that their data can be accessed. It is also vital that the software remains sensitive to the possible situations it might be used in. Above all, it should avoid attracting attention since an adolescent may use it in public situations. For example, an application should avoid playing any audio files without explicit consent from the end user.

   Software design should seek to engage disinterested adolescent clients in particular tasks or subjects. In sensitive situations, it is often inappropriate to enforce use of mobile software. It can be made visually attractive in ways that current therapeutic materials do not afford. Use of animation, colour, and the interaction possibilities the phone affords, may make therapeutic activities more appealing. It may also be possible to give rewards, such as unlocking simple games or some form of feedback based on a client's use of the software. Responses other than default values should be encouraged. This could entail making the interaction sufficiently engaging or enjoyable that it draws the user in.

3. **Ease of use:** Mobile software should require as little training as possible for both clients and therapists. It should also require very little technical support. Clinical settings allow no direct contact between HCI researchers and end users. It is generally only possible to give automated support to adolescent clients, like, for example, the automated sending of forgotten passwords. While it is possible to give more technical support to therapists, it should be a last resort. Therapists are unlikely to use software that requires significant training at the outset or relies on frequent support.

   Because new technological interventions are almost always introduced to clients by therapists, it is essential that they have a clear understanding of them. We recommend presenting any software well in advance to therapists in order to give them time to become comfortable with it. A short introduction session can help to demonstrate the purpose of the software. It should also involve giving detailed suggestions on how therapists might introduce it to a client. Finally, it is invaluable to get therapists trying the software themselves while there is support on hand. Role play is extremely useful. Therapists can take turns playing the role of therapist and playing mock-clients. This helps them practice how they might introduce the software to a client, and also brings up questions and queries that can be dealt with on the spot. Above all, it helps them gain confidence in the technology. Many therapists are particularly comfortable with role-playing as it often is part of their training.

4. **Personal mobile phones:** Many mobile HCI studies evaluate mobile software designed for PDAs or high-end phones that are then given to end users. This is justifiable for many research projects and numerous situations. In this area, we believe it is vital to use clients' own mobile phones. They are a technology in widespread use and represent a practical solution in terms of cost and accessibility. PDAs, while becoming less expensive every year, still carry a significant hardware cost. They require upkeep and technical training and may introduce a novel element into testing that would be difficult to evaluate. Many adolescents will not have used one before and may attract unwanted attention while using one in public. Adolescents already carry their phone wherever they go and are familiar with how it works.

   The decision to use the client's own phone has implications for design and implementation. To achieve necessary coverage, software has to function on a wide range of models, including older models. Designing for such a broad base brings significant restrictions. Almost every device uses different hardware and a different user interface. It is a challenge to design and develop software for all devices in a particular market. The mobile version of java (*JavaME*) enables the researcher to design mobile software that will function and appear as a native application on each different handset. This has the advantage of building on the user's existing knowledge of their device.

5. **Access:** Practical, time and ethical constraints require that any mobile software be accessible to as wide a group as possible. Software should therefore function on a wide variety of handsets, to avoid the potential stigma of it not working on certain older models, and to speed up evaluation. Although increasing access requires more application development time, it should result in quicker evaluation because the software

will be available to a greater number of clients. This means that design is restricted to the limitations of the least advanced phone in the market, unless of course researchers are prepared to develop several versions of the software for different models.

6. **Standard UI and design:** Mobile software should use the standard user interface and interaction style of the user's native phone wherever possible, although a balance has to be struck between this and providing an attractive and engaging interface. Design, graphical and otherwise, should strike a familiar note, with an emphasis on privacy, security, and transparency. Palmblad and Tiplady have made several recommendations for the design of mobile user interfaces for electronic patient reported outcomes (ePRO) that are informative (2004). While their suggestions are directed toward general patient use, particularly on PDAs, many are relevant here. For example, they argue that questions presented to the user should require an active response, and that all attempts should be made to keep content on one mobile screen to require no scrolling. For our purposes, this entails designing for the smallest screen on the market. They also make recommendations regarding text size, colour, and screen layout. It is important to remember that many end users may have learning difficulties, and so, in many cases, text should be kept to a minimum and be as informal as possible.
7. **Error handling:** As with any well-designed application, program errors should be explained practically to the user and without use of technical jargon. Software should handle the burden of any failure, making any necessary operations in the background without requiring any further action from the end user or any further notification. Information loss must be avoided at all costs. Adolescents resent having to repeat their stories to therapists, and are likely to resent having to rerecord any content.

Adaptable Systems

As highlighted earlier, client factors, and the ability of a therapist to work in a client-centred way, are critical factors in the success of mental health interventions. Professionals working in public health-care systems are faced with the need to work with a broad range of clients, from a wide range of sociocultural backgrounds, and experiencing a broad range of disorders. Working in such situations requires therapists to be able to adapt their approaches to suit the needs of a given client. Sociocultural issues are one of the major factors in the development of mental health disorders, and are a key factor that must be considered in treatment. Some of the specific sociocultural issues that affect mental health care include gender, age, social class, sexual orientation, religious or secular assumptions and race, culture, and ethnicity. Systems designed for mental health interventions should ideally reflect this need for adaptation to the needs of clients. Further forms of adaptation are also desirable. For example, teams of therapists working in public health-care settings often have different theoretical backgrounds, and many adopt a variety of eclectic approaches to working with clients.

For technologies to be of practical use in a broad range of MHC interventions, they should ideally be adaptable to (1) a broad range of theoretical models, (2) a broad range of mental health disorders, (3) the differing needs of various demographic groups, and (4) the specific needs of individual clients.

The authors are investigating a form of adaptation that allows therapists to adapt the therapeutic content delivered by systems. For example, some mobile software may include a reminder element where clients receive text message reminders. Different clients may like to receive these at different intervals or at different times. This needs to be easily modifiable. More importantly, the content of these reminder messages needs to be easily adaptable by therapists to the specific needs and motivations of each client. Just as a therapist needs to choose the suitable therapeutic tool for a particular client, they should be easily able to tailor the mobile application.

Adaptable systems offer therapists the flexibility to adjust systems to meet the needs of given clients or client groups. While HCI researchers focus on the design and development of frameworks for delivering therapeutic content, MHC researchers can adapt these systems for use in various interventions and evaluate their therapeutic benefits. Whilst it is desirable that a large number of therapists are comfortable using any system, it is not necessary that all therapists be able to make large scale modifications in the use or therapeutic content of a system. A possible model of use for such a system is:

Small groups of expert MHC researchers and therapists develop and adapt content for use within their own areas of expertise for example, targeting specific disorders and implementing specific intervention models.

Larger groups of therapists use these implementations with clients, and have the option of making small-scale adaptations to suit the needs of specific demographic groups or clients.

By enabling therapists to play an active role in building and tailoring the content delivered by computerized systems, adaptable technologies provide a sustainable approach to the development of client- and therapist-centred technologies for mental health interventions.

Therapists and Designers Work in Parallel

While collaborative design, involving domain experts, is desirable in most design spaces, it has a particular significance in MHC. Difficulties caused by lack of access and ethical constraints are further exacerbated by the fact that few therapists currently have the experience required to design or develop new technologies, or to rigorously evaluate them to the standards required for successful introduction to clinical settings. In order to overcome these limitations, the design and evaluation of mobile applications requires the close collaboration of HCI researchers and therapists.

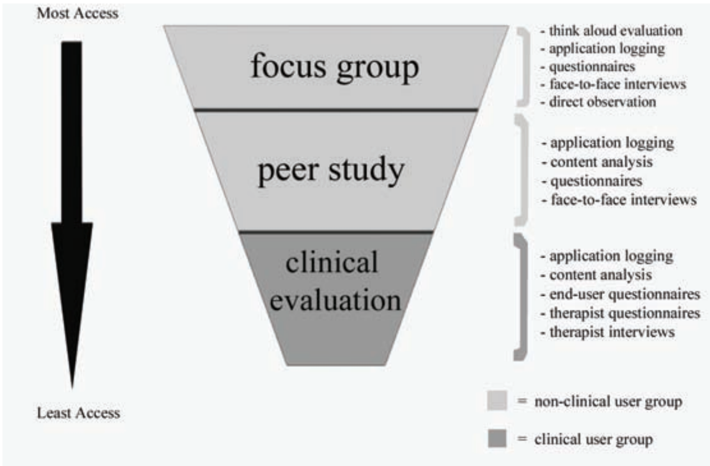
At the outset of the design process, there needs to be a cross-pollination of ideas between the two, and a sharing of knowledge from each expert's domain. The HCI researcher needs to appreciate the sensitivity of the situation, to be able to characterize potential adolescent clients through conversation with the therapist, and to understand a therapist's motivations and typical working practice. The therapist needs to be made aware of the possibilities and benefits afforded by the technology, the things that might not go to plan, and how to present the technology to the client. Therapists must develop protocols that include recommendations for use of technology in practice. Questions of when it is appropriate to use a piece of software, how long it should be used for, and how it should be incorporated into the therapeutic process, are questions that lie in the domain of mental health research. Hence, a parallel cycle of development of protocols can be carried out by mental health therapists, while still feeding back information of relevance to the design and usability of the application.

The HCI researcher and therapist both contribute to identifying needs or areas where technology can add to the situation, and to developing and appraising the initial application design. Characterization of end users can proceed through collaboration with therapists and detailed analysis of the relevant literature. But, it is clear that some questions are only answerable by end users. Both parties are interested in ensuring the best outcome for the client, but where the therapist is focused on the overall effectiveness of the intervention, the HCI researcher is concerned with designing the most effective software.

Multistage Prototyping

Ethical and access restrictions make iterative design and user-centred design problematic. In order to overcome these limitations and to maximise evaluation possibilities, we advocate a three-stage approach for evaluating mobile software (see Figure 1). In the first two stages, these groups consist of adolescents from similar backgrounds to the targeted end users, but who do not have mental health problems. The most detailed HCI evaluation takes place in the early stages of this process while there are fewer limitations to the methodologies

Figure 1. Three stage evaluation



available. At each subsequent stage, the evaluation and data collection possibilities become restricted as access and ethical limitations increase. The main goal of this section is to describe a process that refines design through the use of adolescent peer groups.

Stage One: Focus Group

As described earlier, a close working relationship between the HCI researcher and therapist is essential as a foundation to the design process at this stage. At this stage, once a particular design has been identified, several paper prototypes are created. Paper prototypes are particularly useful in providing initial common ground for the two parties, who are, at the outset, approaching the design from very different perspectives. They help open a dialogue between the therapist and the HCI researcher around different designs, leading to more focused and detailed designs. These designs are initially tested informally with colleagues from HCI and mental health backgrounds. A working prototype is developed once a particular design has been chosen. Next, a pilot group of between 5 and 10 adolescents is formed. These adolescents do not have mental health problems, but are chosen based on similar socioeconomic, educational, and geographical backgrounds to the ultimate adolescent end users. The prototype is loaded onto participants' own handsets, and detailed qualitative data can be gathered through face-to-face interaction with researchers. Evaluation at this stage focuses on usability and the overall appeal of the software. Evaluation methods available at this stage include think aloud evaluation, application logging, questionnaires, interviews, and direct observation in a controlled environment.

Stage Two: Peer Group User Study

The data collected from stage one is collated and analyzed. It is used to inform any changes to the original prototype. Substantial design changes can still be made at this point. Once these changes have been implemented, the software is tested on a larger group of adolescent peers (typically 20 to 50), again, from similar backgrounds to the intended end users. This stage aims to evaluate the design across a wider group of adolescents, with sufficient numbers to obtain statistically significant results.

Schools are the most convenient partners to obtain access to large groups of adolescents. Testing can take place over a number of weeks, depending on the requirements of the evaluation. Whether training or technical support are provided again depends on the mobile software being evaluated and the aim of the study. A short group introduction should be given to participants explaining the purpose of the testing. It should be made clear that it is the software rather than the participant that is being tested. The software is then loaded onto the phones of participating adolescents. Data collection at this stage depends on what exactly is being evaluated, and can involve qualitative and quantitative measures. There is a focus on usability across a wide range of users and handsets, as well as on content created and emotional engagement with the software. Measures appropriate include application logging, analysis of any content created if relevant to the mobile software being tested, questionnaires, and face-to-face interviews. Access and ethical restrictions are still relatively relaxed at this stage, but the number of participants inhibits extensive qualitative data collection.

Stage Three: Clinical End-User Evaluation

This is the final stage of evaluation. Research must not proceed to this stage unless there is confidence in the outcomes. It begins by analyzing all the data from the previous stage and implementing changes into the final prototype. This stage of testing involves clinical trials with adolescents who are attending therapists. The number of participants will vary according to research resources and therapist and client availability. The authors recommend 15 to 20 as a manageable number. Evaluation proceeds in conjunction with the participating therapists. Available data collection measures include application logging, content analysis, succinct end-user questionnaires, therapist questionnaires, and interviews. Analysis of the content collected is particularly important. Questionnaires are possible at the end of testing, but should be short, and need to address some therapeutic elements. Short interviews may be possible via proxy, but will include a therapeutic perspective. More detailed feedback can be collected from the therapist via questionnaire and interviews, covering the impression the technology had on the client.

Once these three stages have been completed, it is possible to implement final changes and make the software generally available to therapists. The objective of this design approach is to reduce the burden of evaluating the mobile software for therapists and end users to a minimum, as well as reducing the time taken to iterate through design approaches with adolescent peers and increase the amount of feedback on the design. What follows is a case study giving a specific example of this process in action. This study serves not to present any particular results, but rather as a holistic example of the methodology in practice. It also serves to demonstrate some of the recommendations discussed earlier.

CASE STUDY: "MOBILE MOOD CHARTING"

As an illustration of the principles and methodologies discussed, we describe the design and evaluation of one particular mobile application called Mobile Mood Diary. It is a mobile-phone version of a paper diary that therapists often give to clients to monitor mood. The main significance of this case study, as presented, is as a demonstration of

Figure 2. Fragment of paper diary

Date &Time	Energy How much energy do you have out of ten? (please put an 'X' on the scale)	Sleep How many hours did you sleep last night?	Mood How do you feel today out of ten? (please put an 'X' on the scale)	Diary – please describe how you feel today.
1/05/06 10.15	1 2 3 4 5 6 7 8 X 10 very low very high	I slept 8 hours.	1 2 3 4 5 6 7 8 9 X very low very happy	I feel great today, loads of energy.
	1 2 3 4 5 6 7 8 9 10 very low very high	I slept ____hours.	1 2 3 4 5 6 7 8 9 10 very low very happy	

the multistage prototyping process. We will not go into specific analysis of individual evaluation of elements of the software, although we will touch upon them in passing. This section is focussed on discussion of the practicalities of putting this methodology into practice.

"Mood monitoring" is an important component of cognitive behavioural therapy (CBT). This involves adolescents recording their mood at regular intervals, to help them to recognize the factors that may be impacting on their mood, and how they have managed at times when they have felt depressed. "Mood monitoring" is traditionally done as a pen and paper exercise that adolescents are asked to complete on a daily basis, using a mood chart supplied by the therapist. Young people are notoriously poor at completing these charts or remembering to bring them to the next session (Feldman Barrett & Barrett, 2001; Lester, 1999). This case study describes a comparative evaluation of a mobile mood diary and a paper mood diary (see Figure 2).

Design of the Mobile Chart

The initial design of the mood diary involved one dedicated HCI researcher and one therapist working together in consultation with several other HCI experts and therapists. The main goal of the software was to enable adolescents to record their mood (as well as energy, hours of sleep, and diary recording) in a streamlined process. Recording a mood needed to be a series of clear and familiar actions that could be accomplished quickly by adolescent end users. The objective was to make the software as intuitive to use as a paper diary.

In accordance with the design recommendation to provide wide access, the software had to work on the majority of adolescents' own phones. From a technical viewpoint, it was important to appraise different solutions to achieving this. The transfer of data is possible via IR, Bluetooth, or HTTP; however, the criterion of access ruled the first two out, as current phones in the adolescent market did not provide sufficient support. Furthermore, both technologies would have required that the information be transferred during a therapy session, potentially taking up valuable time**.** It is also possible that there might have been technical hitches that could have undermined the client's confidence in the therapist.

An initial survey of current phones on the market over the past year revealed that there was stronger support for *JavaME* than other software platforms. Further surveys at the first two stages of evaluation supported this, revealing an even wider support base. This meant that clinical evaluation could proceed with the confidence that the

Figure 3. Screens from final mobile diary design

majority of client devices would be supported by the software. *JavaME* provided a reliable method of providing an interface that behaves and looks like the native user interface of each individual phone, as well as a robust way of reaching a wide number of existing handsets.

Stage One: Focus Group User Study

The initial prototype of the mobile mood diary included sliding bars for recording mood and energy (both out of 10), as well as hours slept, and a space for texting thoughts. A single mobile screen was used for all screen elements. Initial testing took place with six 12 to 14 year olds in an inner-city youth club, and was in line with the wide range of activities carried out there. The application was installed on their phones and they were asked to complete a series of simple tasks, while following the think-aloud protocol. They were then allowed 30 minutes free use of the software and then interviewed for general opinions. They also completed questionnaires.

Significant usability issues with the initial prototype were identified at this stage. Some participants failed to notice some screen elements and became frustrated with the interaction of selecting different screen objects. On some handsets this required vertical scrolling to navigate the content. As a result of the initial evaluation, the prototype was changed to include four distinct pages, one for each of the elements to be recorded (see Figure 3). Other modifications included making the user interface more graphically appealing, replacing the default standard scale with a game metaphor of a moving character along a 1 to 10 scale, and streamlining the navigation structure.

The purpose of this stage was to reveal any significant interaction difficulties and get detailed usability feedback from adolescent users. This evaluation stage allowed for hands-on involvement from designers and the opportunity for participants to actively contribute to the design process. By the end of this stage, there was a clear idea of what worked and what needed to be changed.

Stage Two: Peer Study

Following significant redesign, a second evaluation in three different schools was carried out. A nonclinical sample of 70 self-selecting students took part in the study with school and parental permission. Participants' ages ranged from 13 to 17 (n=73 $mean$=14.87 SD=1.141) and were mostly female (86.3%). Fifty-two students were given paper-based charts and 21 students had the mood diary loaded onto their phones. The purpose of this stage was to evaluate compliance and usability of both the paper and mobile diary. Due to the emphasis on entering real-life data (i.e., moods) and the more formal setting, a greater overhead was involved in securing access and parental consent, and care had to be taken regarding privacy. The experiment proposal was reviewed by an ethics board in the same fashion as a clinical evaluation. While a full description of this study is beyond the scope of this chapter, we present the format and main parameters, along with a summary of the results.

Testing took place over a 2-week period. Both groups were asked to record their mood once a day, although they were not reminded to do so. No incentive was given to either group. Data collection at this stage included data logging, analysis of the recorded content, and postexperiment questionnaires. Posttrial face-to-face interviews were also used with both groups to gauge general opinions, overall satisfaction, and suggested improvements.

In order to assess whether a paper vs. mobile phone-based platform influenced compliance behaviours, participants who completed paper based diaries ($n=16$) and participants who completed a mobile phone-based diary ($n=17$) were assigned to two groups. Compliance was judged as completing at least one entry per day. Data was analysed using an independent samples t-test in order to test for significant differences in the number of mood entries participants completed. Groups were found to differ significantly in their rates of actual compliance ($t= -2.324, p< .027$), with the mobile phone-based diary group ($mean=8.12$) producing significantly more entries than the paper-based diary group ($mean=5.44$).

The study also provided qualitative data that indicated that the adolescents who were using the mobile phone were happier with the level of privacy the phone provided, and found it as easy to use as the paper method and significantly more appealing. Feedback strongly suggested that adolescents valued security and suggested that a security feature (e.g., a pin) be added to the software. Overall, it was a successful evaluation of the mobile mood diary that did not require a technical demonstration or technical support during the testing period. The strength of the results provided ethical justification for proceeding to the third stage of the evaluation, firstly as there clearly seemed to be a benefit associated with the software, and secondly that no technical or major usability issues arose over the 2-week period of each evaluation.

Stage Three: Clinical End-User Evaluation

By the clinical stage, the software had been evaluated with a large number of adolescents. The analysis of feedback received from the school pilot study was used to make modifications to the mobile software. Additions suggested from stage two (e.g., a security pin) were added. Beyond the software itself, it became clear that it is important to identify clients' mobile phones in advance in order to determine whether they supported java, in order to avoid disappointing adolescent end users.

Clinical testing of the mobile mood diary is ongoing and is being used as a means to refine the protocol for final use of the software. A nationwide clinic is currently evaluating the software. Several regional clinics and individual therapists are participating as well in order to provide evaluation results across different practices. While basic usability issues are not expected to be discovered at this stage of evaluation, past experience of clinical trials in the mental health area leads us to expect that issues related to the use of the content within sessions will emerge in the evaluation and feedback into subsequent versions of the diary.

Discussion

A significant advantage of the two first stages of evaluation was that the procedures, protocols, and data analysis techniques were pinned down before clinical testing began. The school study raised a number of practical issues. For example, it became clear that it was important to develop fallback options in case the software did not work on participants' phones, which could result in them feeling left out. This is particularly important regarding mobile phones, which have considerable social value for adolescents. During the initial installation of the software on phones, the software failed to work on the phones of two participants. Their reaction was that their phone was "not good enough." Experimenters reassured participants that whether the software worked on their phones had nothing to do with how "good" their phones were. It is important to avoid this stigma. For clinical usage, it may be useful to develop a simple tool that would allow therapists to enter the model of the client's phone to find out if the software is compatible with it.

It became clear, after the first two stages, that the manner in which the evaluation study was presented to both therapists and clients was critical. For example, when introducing the software to therapists interested in participating in trials, it was important to make it clear that we were not evaluating a system for two-way communication with clients. The school evaluation revealed trust issues relating to how the software was described to adolescents. At the outset, experimenters told participants they would be "installing some software" on their mobile phones. This phrase made some adolescents suspicious of the testing, and potentially afraid they were being screened for mental health problems or that the software might "damage" their phones. Describing the installation as a "download," like a game or a ring tone rather than a program or an application, was sufficient to allay such fears. The context and type of language used to describe new technologies to adolescents is important; protocols that deal with these language and context issues would help experimenters and therapists when they are introducing new technologies to adolescents. In general, the importance of how the software is introduced to all participants (e.g., to schools, parents, therapists, adolescent peers, adolescents) cannot be overvalued.

Over the course of testing, privacy and security were concerns for many of the participants using paper diaries. Paper methods have drawbacks, such as being easy to lose or damage. They do not provide sufficient privacy to the end user. Mobile phones can handle private information in a more unobtrusive fashion. The end user needs to know exactly what the protocol for dealing with their information is, how much control they have over it, and whether anyone else has access to it.

The school study provided good statistical data for the appeal and effectiveness of the mobile software. This was an important element in convincing therapists (often initially sceptical) to participate in evaluation. It also provides confidence in the likely outcomes at the clinical stage of evaluation.

The case study provides an illustration of the multistage prototyping methodology in action. This process provides the designer with a range of evaluation data that was not possible to obtain through other methods. Furthermore, it brought up significant difficulties with the software and protocols in safe situations before introducing a modified version into a sensitive situation. Usability, appeal, and protocols were all tested in situations where there was time and the possibility of failure. This enabled the software design to be refined into one that could be used with confidence in this sensitive situation.

CONCLUSION

Mobile phone-based software holds significant promise for engaging adolescents in therapeutic activities, but designing such applications is extremely challenging due to the ethical and access constraints that curtail access to end users. This chapter has examined the design and evaluation approaches that may be used in these situations. We began by examining the characteristics and requirements that sets this area apart from other application areas, although relevant research from education and medical health care is considered. The emerging role of mobile technologies in therapeutic practice was explored, along with the identification of specific activities where it can have the most beneficial effect.

There are distinct barriers and limitations faced by researchers when designing software for mental health. These constraints render standard approaches to user-centred design impractical and potentially unethical. To address these issues, a set of design recommendations has been presented. Designs should emphasize engagement, privacy, and data security for the end user. There is a need to work closely with therapists, and there are also benefits to developing systems that are easily adapted and modified by therapists. We have also discussed the typical measures available for data gathering for the evaluation of mobile applications and how suitable they are in this field.

We have presented a three-stage prototyping process for the design and evaluation of mobile software for adolescents. The mobile mood diary case study provides an illustration of how the design recommendations, and particularly the multistage prototyping process, may operate in practice. Conducting initial evaluations with focus groups and then peer groups permits researchers to improve the software and to proceed to clinical evaluations with confidence in the outcomes.

Research on the design of technology within the area of mental health is still at an early stage. There is a need to establish which therapeutic tasks are best suited to mobility and how to design for them. At the moment, software is designed to compliment current therapeutic practice; in the future it may change practice to a degree. For the researcher, it can be difficult working in constrained and restricted circumstances, but positive results can have a dramatic impact on people's lives.

REFERENCES

Assay, T. P., & Lambert, M. J. (1999). The empirical case for common factors in therapy: Quantitative findings. In B. L. Duncan, M. L. Hubble, & S. D. Miller (Eds.), *The heart and soul of change* (pp. 23-55). Washington, DC: American Psychological Association.

Barendregt, W., & Bekker, M. M. (2005). Extended guidelines for usability (and fun) testing with children. In *Proceedings of SIGCHI.nl Conference 2005.* Den Haag.

Bauer, S., Percevic, R., Okon, E., Meerman, R., & Kordy, H. (2003). Use of text messaging in the aftercare of patients with bulimia nervosa. *European Eating Disorders Review, 11*, 279-290.

BBC. (2006). *Youngsters reliant on mobiles*. Retrieved 5/3/2007, from http://news.bbc.co.uk/2/hi/business/5358784.stm

BMA. (2006). *Child and adolescent mental health—A guide for healthcare professionals.* London: Board of Science of the British Medical Association.

Branch, J. L. (2001). Junior high students and think alouds Generating information-seeking process data using concurrent verbal protocols. *Library & Information Science Research, 23*, 107-122.

Brukman, A., & Bandlow, A. (2002). HCI for kids. In J. J. a. A. Sears. (Ed.), *Handbook of human-computer interaction*. NJ: Lawrence Erlbaum Associates.

Carroll, J., Howard, S., Peck, J., & Murphy, J. (2002). A field study of perceptions and use of mobile telephones by 16 to 22 year olds. *Journal of Information Technology Theory and Application*.

Caspar, F. (2004). Technological developments and applications in clinical psychology: Introduction. *Journal of Clinical Psychiatry, 60*(3), 221-238.

Coyle, D., Doherty, G., Matthews, M., & Sharry, J. (2007). Computers in talk-based mental health interventions. *Interacting with Computers.* In press, DOI: 10.1016/j.intcom.2007.02.001.

Coyle, D., Matthews, M., Sharry, J., Nisbet, A., & Doherty, G. (2005). Personal investigator: A therapeutic 3D Game for Adolescent Psychotherapy. *International Journal of Interactive Technology and Smart Education, 2,* 73-88.

Druin, A. (2002). The role of children in the design of new technology. *Behaviour and Information Technology, 21*, 1-25.

Egan, G. (2002). *The skilled helper: A problem-management and opportunity-development approach to helping* (7th ed.). Pacific Grove, CA: Wadsworth Publishing.

Feldman Barrett, L., & Barrett, D. J. (2001). An introduction to computerized experience sampling in psychology. *Social Science Computer Review, 19*, 175-185.

Gosbee, J., & Ritchie, E. (1997). Human-computer interaction and medical software development. *Interactions, 4*(4), 13-18.

Grinter, R. E., & Eldridge, M. (2001). *Y do tngrs lv 2 txt msg? Why "texting" became popular with teenagers.* . Paper presented at the Proceedings of the Seventh European Conference on Computer Supported Cooperative Work (ECSCW) September 18-20 Bonn, Germany.

Hanna, L., Risden, K., & Alexander, K. J. (1997). Guidelines for usability testing with children. *Interactions, 4*(5), 9–14.

Hanna, L., Risden, K., Czerwinski, M., & Alexander, K. (1999). The role of usability research in designing children's computer products. In A. Druin (Ed.), *The design of children's technology* (pp. 3-26). San Francisco: Morgan Kaufman.

Lester, J. S., B. & Stelling, G. (1999). Lifelike pedagogical agents for mixed-initiative problem solving in constructivist learning environments. *User Modeling and User-Adapted Interaction, 9*(1-2), 1-44.

Leysia, P., & Marilyn, S. (2002). Voice-mail diary studies for naturalistic data capture under mobile conditions. In *Proceedings of the 2002 ACM conference on Computer supported cooperative work*. New Orleans, LA: ACM Press.

Markopoulos, P., & Bekker, M. (2003). On the assessment of usability testing methods for children. *Interacting with Computers, 15*(2), 227-243.

Murray, C. L., & Lopez, A. D. (Eds.). (1996). *The global burden of disease: A comprehensive assessment of mortality and disability from disease, injuries, and risk factors in 1990 and projected to 2020*. Cambridge, MA: Harvard University.

Naismith, L., Lonsdale, P., Vavoula, G., & Sharples, M. (2004). *Mobile technologies and learning (report commissioned by NESTA FutureLab)*. Birmingham: University of Birmingham.

Nakamoto, H., Kawamoto, A., Tanabe, Y., Nakagawa, Y., Nishida, E., Akiba, T., & Suzuki, H. (2003). Telemedicine system using a cellular telephone for continuous ambulatory peritoneal dialysis patients. *Advances in Peritoneal Dialysis, 19*, 124-129.

Neville, R., Greene, A., McLeod, J., Tracy, A., & Surie, J. (2002). Mobile phone text messaging can help young people manage asthma. *British Medical Journal, 325*.

Norton, M., Wonderlich, S. A., Myers, T., Mitchell, J. E., & Crosby, R. D. (2003). The use of palmtop computers in the treatment of bulimia nervosa. *European Eating Disorders Review, 11*(3), 231 - 242.

Offer, D., Howard, K. I., Schonert, K. A., & Ostrov, E. (1991). To whom do adolescents turn for help? Differences between disturbed and nondisturbed adolescents. *Journal of the American Academy of Child and Adolescent Psychiatry, 30*(4), 623-630.

Office for National Statistics. (2005). *Mental health in children and young people in Great Britain, 2004*. London: HMSO.

Office of the Deputy Prime Minister. (2004). *Mental health and social exclusion*. Social Exclusion Unit report. London: ODPM.

Palmblad, M., & Tiplady, B. (2004). Electronic diaries and questionnaires: Designing user interfaces that are easy for all patients to use. *Quality of Life Research, 13*, 1199-1207.

Roberts, L., & Dyer, A. (Eds.). (2004). *Concise guide to ethics in mental health care*. Washington, DC: American Psychiatric Publishing.

Roschelle, J., Sharples, M., & Chan, T. W. (2005). Introduction to the special issue on wireless and mobile technologies in education. *Journal of Computer Assisted Learning, 21*(3), 159-161.

Royal College of Psychiatrists. (2004). *Mental health and growing up (3e). Factsheets for parents, teachers and young people*. London: Gaskell.

Sanderson, K., Andrews, G., Corry, J., & Lapsley, H. (2003). Reducing the burden of affective disorders: Is evidence-based health care affordable? *Journal of Affective Disorders, 77*(2), 109-125.

Savill-Smith, C. (2005). The use of palmtop computers for learning: A review of the literature. *British Journal of Educational Technology, 36*(3), 567-568.

Schaefer, C. E., & Reid, S. E. (Eds.). (2001). *Game play: Therapeutic use of childhood games*. New York: Chichester: Wiley.

Suchman, L. A. (1987). Plans and situated actions: The problem of human-machine communications. Cambridge, UK: Cambridge University Press.

Surgeon General. (1999) *Mental health: A report of the Surgeon General—Executive Summay*. Washington DC, Dept. of Health Human Services. Retrieved August 2006, from http://www.surgeongeneral.gov/library/mentalhealth/home.html

Vavoula, G., Meek, J., Sharples, M., Lonsdale, P., & Rudman, P. (2006). *A lifecycle approach to evaluating MyArtSpace*. Paper presented at the Wireless, Mobile and Ubiquitous Technologies in Education (WMUTE) 2006.

Vavoula, G., Sharples, M., Rudman, P., Lonsdale, P., & Meek, J. (2006). Learning bridges: A role for mobile technologies in education. *Educational Technology Magazine, special issue on mobile learning*.

Wood, J., Keen, A., Basu, N., & Robertshaw, S. (2003). *The development of mobile applications for patient education*. San Francisco: ACM Press.

KEY TERMS

Client: The person undergoing therapy. Therapists typically differentiate between three types of clients: adults, adolescents, and children. In this chapter, client generally refers to an adolescent client.

CBT: Cognitive behavioural therapy is a highly structured psychotherapeutic method used to alter distorted attitudes in order to change behaviour and emotional state.

JavaME: The microedition of java programming language used for developing software on mobile phones.

Mental Health: Can be defined as successful performance of mental function that results in productive activities, fulfilling relationships, the ability to adapt to change, and the ability to cope with adversity.

Mental Health Care (MHC): Aims to help people improve their psychological well-being.

Mental Illness: can be defined as the experience of psychological symptoms that are severe enough that normal functioning is impaired and help is needed to recover.

Mood Charting: A form of diary study that is used in some therapeutic practices to monitor changes in mood.

Multistage Prototyping: This is a process the authors have applied to the evaluation of mobile software to be used in mental health situations with adolescents.

Psychotherapy: The treatment of a behaviour disorder, mental illness, or any other condition by psychological means.

Sensitive Situation: A situation where poor-quality solutions may have a highly negative impact, and where the introduction of not just the technology, but also the designer, could be detrimental to the environment that is the target of the technological intervention.

Chapter XXXIX
Widely Usable User Interfaces on Mobile Devices with RFID

Francesco Bellotti
University of Genoa, Italy

Riccardo Berta
University of Genoa, Italy

Alessandro De Gloria
University of Genoa, Italy

Massimiliano Margarone
University of Genoa, Italy

ABSTRACT

Diffusion of radio frequency identification (RFID) promises to boost the added value of assistive technologies for mobile users. Visually impaired people may benefit from RFID-based applications that support users in maintaining "spatial orientation" (Mann, 2004) through provision of information on where they are, and a description of what lies in their surroundings. To investigate this issue, we have integrated our development tool for mobile device, (namely: MADE, Bellotti, Berta, De Gloria, & Margarone, 2003), with a complete support for RFID tag detection, and implemented an RFID-enabled location-aware tour-guide. We have evaluated the guide in an ecological context (fully operational application, real users, real context of use (Abowd & Mynatt, 2000)) during the EuroFlora 2006 international exhibition (EuroFlora). In this chapter, we describe the MADE enhancement to support RFID-based applications, present the main concepts of the interaction modalities we have designed in order to support visually impaired users, and discuss results from our field experience.

INTRODUCTION

Starting from the European Union cofounded E-Tour project, we designed the tourist digital assistant (TDA) concept and developed multimedia tour guides on mobile devices (PocketPC and Smartphone devices) for a number of European tourist sites, such as the Costa Aquarium of Genoa, "Strada Nuova" architectonical area and the city of Genoa, the Castellon region in Spain, and the city of Uddevalla in Sweden (Bellotti, Berta, De Gloria, & Margarone,, 2002).

The tour guide provides multimedia contents, added-value information, and location-based services to the tourists. Added-value services are implemented by integrating the mobile devices with additional hardware and software tools such as GPS, electronic compasses, wireless connectivity, digital cameras, written text input, databases, and so forth.

See Figure 1 for snapshots of tourist guide applications.

Relying on the argument that "play is a powerful mediator for learning throughout a person's life," we developed the "educational territorial-gaming" concept in VeGame (Bellotti, Berta, De Gloria, Ferretti, & Margarone, 2003), a computer-supported educational wireless team-game played along Venice's narrow streets to discover the art and the history of the city (see Figure 2), and in ScienceGame (Bellotti, Berta, De Gloria, Ferretti, & Margarone, 2004), a sort of treasure-hunt game inviting players to discover the mysteries and the marvels of the science (see Figure 3) during the "Festival della Scienza" exhibition held in Genoa every year.

These applications were developed from scratch. From these first experiences, we identified common needs and came up with a system to support design of multimedia applications for mobile devices, called Mobile Applications Development Environment (MADE) (Bellotti et al., 2002).

MADE includes M3P (MicroMultiMedia Player), a network-enabled multimedia player easily programmable through the micromultimedia services language (MSL). MSL provides high-level components encapsulating advanced services (e.g., positioning, database query, path search, etc.) that can be easily integrated in multimedia applications. This allows building modular software programs that provide information-rich services to the general public through a coherent and homogeneous HCI that can be learned with low mental workload. On the other hand, MADE hides the low-level aspects of multimedia and service management, allowing designers to focus on the modalities of presentation of information and on user interaction, reducing learning, development, and code maintenance time.

In this chapter, we describe the latest MADE enhancement: we have integrated it with a complete support for RFID detection to allow development of multimedia mobile applications directly connected with the physical world (Want, Fishkin, Gujar, & Harrison, 1999). All low-level aspects of the hardware tag-detection system that are neces-

Figure 1. Snapshots from the Aquarium and Strada Nuova tour guides on PocketPC device

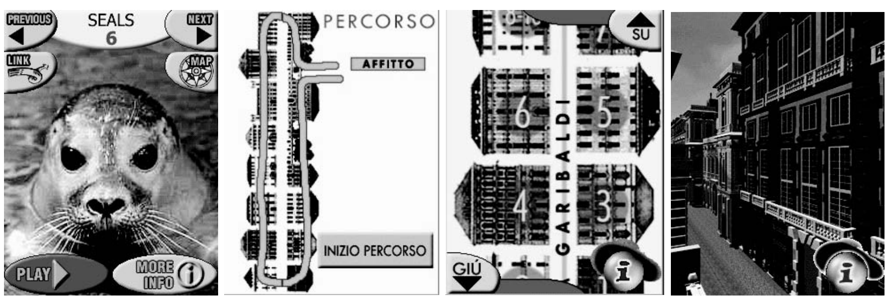
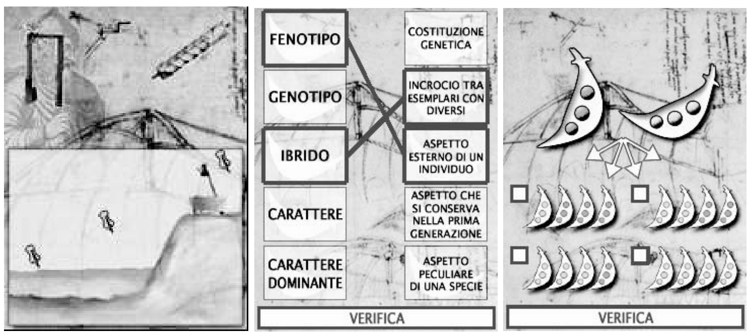


Figure 2. Snapshots from VeGame

Figure 3. Snapshots from ScienceGame



sary to identify and locate physical objects with attached small RF tags (Want, 2004) are hidden to MSL programmer by the MADE system.

This chapter will also show the use of MADE with the RFID support in a real context such as EuroFlora 2006 international exhibition. This guide differs from others because it has been ad-hoc developed in order to meet strict usability needs. In particular, the novel interface design assists visually impaired people in maintaining "spatial orientation" (Mann, 2004) through provision of information on where they are, hazards that might be in the way, and a description of what lies in their surroundings.

MADE SUPPORT OF RFID TECHNOLOGY

Location-Aware Computing

Recent research has developed several systems, to determinate physical location, that differ by accuracy, cost, and coverage (Boriello, Chalmers, La Marca, & Nixon, 2005). The global positioning system (GPS), which uses signal from satellite to estimate position (Djuknic & Richton, 2001), is the most used system, but only for applications in outdoor areas. In indoor areas and urban areas with poor sky visibility, the system does not work properly. Moreover, it has a long start-up time.

To overcome these limitations, the first indoor positioning system was the active badge system (Want, Hopper, Falcão, & Gibbons, 1992), which is based on sensors that receive infrared ID broadcast from tags worn by people. This system gives a poor (room-grained) localization precision. After the active badge system, typical indoor location systems are based on radio frequency and on the estimation of position computed from the measured signal strength. Various technologies can be used: Wi-Fi (Howard, Siddiqi, & Sukhatme, 2003), Bluetooth (Bruno & Delmastro, 2003) and nowadays RFID (Liu, Corner, & Shenoy, 2006).

The first two solutions can give an accuracy of around some meters, but require expensive fixed base stations. RFID tags, instead, are very inexpensive and have the same performance. The literature reports also of many location estimation algorithms based on cellular radio networks (Xu & Jacobsen, 2005). However, there is not a generally agreed solution today, and each algorithm has pros and cons, depending on environmental issues. Finally, some vision-based algorithms (López de Ipiña, Mendonça, & Hopper, 2002) are promising because they do not require infrastructure (like tags, satellite, or base station). However, it is difficult to set up a system to locate a user with a 1-meter precision. In the selection of the best methodology for our system, we have taken into account three major issues: the possibility to have a system for outdoor/indoor spaces (like the EuroFlora 2006 exhibition area), a technology with a low cost for the deployment of the infrastructure, and a likely pervasive availability of the system in the near future. All these requirements are satisfied by the RFID technology.

RFID Application Fields

Major RFID application domains include monitoring physical parameters, such as temperature or acceleration, during fragile or sensitive products delivery, monitoring product integrity from factory to retail locations (Siegemund & Floerkemeier, 2003), utilities for home and office automation (Langheinrich, Mattern, Romer & Vogt,, 2000). Nowadays we have passive or active inexpensive RFID (approaching 35 cents today, with a goal of 5 cents (Quaadgras, 2005)) that makes these kinds of sensors practical for tourist applications. For example, a museum exposition can place tags attached to each point of interest so that tourists can receive information about exposition in the right moment at the right place; when near to the object. The research community has actively explored this possibility at the Exploratorium, the interactive science museum in San Francisco. The HP Laboratories researchers have implemented a system that uses three types of identification technology: infrared beacon, barcodes, and RFIDs (Fleck, Frid, Kindberg, O'Brian-Strain, Rajani, & Spasojevic, 2002). In Goker et al. (Goker, Watt, Myrhaug Whitehead, Yakici, Bierig, et al., 2004), a special tag that can work with mobile devices to provide ambient information to users on the move is described. In the Cooltown project (Kindberg & Barton, 2001), RFIDs are used to attach pointers from everyday objects to entities in the computational world. A full exploitation of RFID potentials requires study and implementation of human–computer interaction (HCI) modalities able to support usability of the enhanced mobile tool by the general public. This implies the necessity to resort to programming methodologies and tools specifically dedicated to support the RFID technology. Thus, we have extended the MADE toolkit to support a link between applications and physical world through RFID sensors.

MADE Architecture

A typical MADE application consists of a set of pages containing multimedia and service objects. The micromultimedia services language (MSL) script specifies pages' layout and objects' appearance, synchronization, and user-interaction modalities. MSL scripts are interpreted at runtime by the M3P player that manages presentation of contents and user interaction according to the instructions specified in the input MSL script.

M3P player relies on two-layer architecture (see Figure 4) involving a high-level, platform-independent director and a low-level driver. The director is responsible for creating, initializing, and managing the objects that implement the language functionalities. In order to support incremental development of the player, M3P is composed by a set of modules. In particular, the director has been designed to be independent of the components it manages. According to the object-oriented methodology, this has been achieved by encapsulating the functions of the components in the code of their class, and by structuring the hierarchy so that the director can simply keep a reference to the presentation's pages and convey them events.

According to the instructions specified by the MSL developer in the script, events (either from the system or from user interaction) are conveyed to the director that simply redirects them to the components of the page currently on show, which is the higher-priority choice or, with lower priority, to the other pages of the presentation.

Events are implemented as string messages that are to be interpreted by the target objects. This design choice allows the director's code to be independent of the components and the components to be independent of each other. The basic assumption of this schema is that each component exports a well-defined interface (i.e., a set of messages to which it is able to react) and implements this interface (i.e., implements the reaction to such messages).

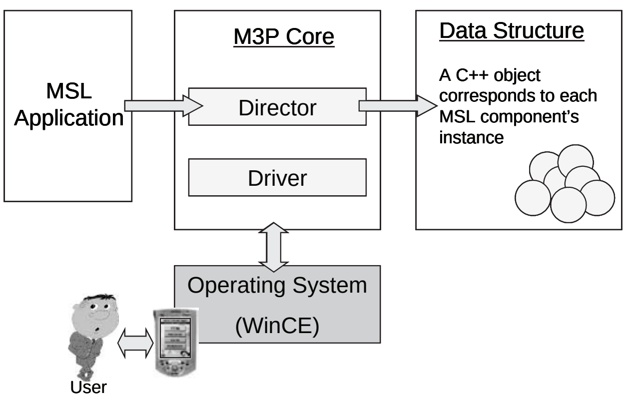
Thus, components can be seamlessly added and interchanged (in this last case, as long as they comply with the same interface). Adding a new component (i.e., a new functionality) does not involve any change either in the director's code, or in the other components' code.

Such a design choice supports easy incremental development, allowing seamless integration of services within a single application framework. This implies that a homogeneous HCI can be applied to an application that hosts several different services that have been developed independently of each other (e.g., intelligent tour planning, interactive maps, positioning, and database access).

MSL relies on a component-based data structure. That is, an MSL file specifies creation of components, attributes of components, and their reaction as a consequence of user interaction. Components are organized in three main libraries: multimedia (e.g., audio, image, video, button), synchronization (utilities like timers that can be used to implement synchronization and scheduling of contents), and services (objects that encapsulate services such as positioning, shortest path search, tour planning, database query, etc).

Every different type of component has its own kind of attributes (fields). The fields record data for specifying the appearance (such as position) and the behaviour (i.e., reactions to events) of the component. In general, components are contained in a special container component, called CARD, that can be thought of as an empty page on which the developer can add components.

Figure 4. MADE architecture



The core of M3P involves a platform-independent director and a platform-dependent driver. The director manages the multimedia objects that implement the presentation. Objects are compounded in hierarchical structures. For instance, a CARD (i.e., a multimedia page) may include several images, buttons, and mpeg players. The driver implements the functions to access the hardware, while the director deals with the logic of the multimedia presentation.

Integration of RFID Subsystem

A major feature of MADE consists in the possibility of incrementally adding new hardware and software modules, that are integrated into the HCI framework with no need for modifying the M3P core, since every component's class is responsible for interpreting its receivable messages, independent of the others. MADE can integrate, into a common framework, various hardware modules independently developed to augment the mobile device potentiality. M3P driver's classes, which have to be developed to integrate every new hardware subsystem, manage low-level aspects of the hardware modules, while the MSL interface to the application developer abstracts the services at high level. This implies that a homogeneous HCI can be applied to an application that hosts several different services that have been developed independently of each other (e.g., automatic positioning, intelligent tour planning, and database access can be integrated in an interactive map), and the MSL developer can simply exploit the service modules focusing on the integration of the HCI. Examples of hardware modules already integrated in MADE are a positioning and orientation module that an MSL developer can exploit to get geographical position from a GPS receiver and direction from a digital compass, and the remote communication module able to exploit the hardware available for connection with the external world (e.g., wired/wireless LAN, Bluetooth, GSM/GPRS cellular networks).

In order to enable applications to react to objects in the physical world, our new M3P module, called RFID sensing module (RfidSM), detects presence of RFID tags in the surrounding space and notifies the M3P run-time objects with events implemented as string messages (see Figure 5).

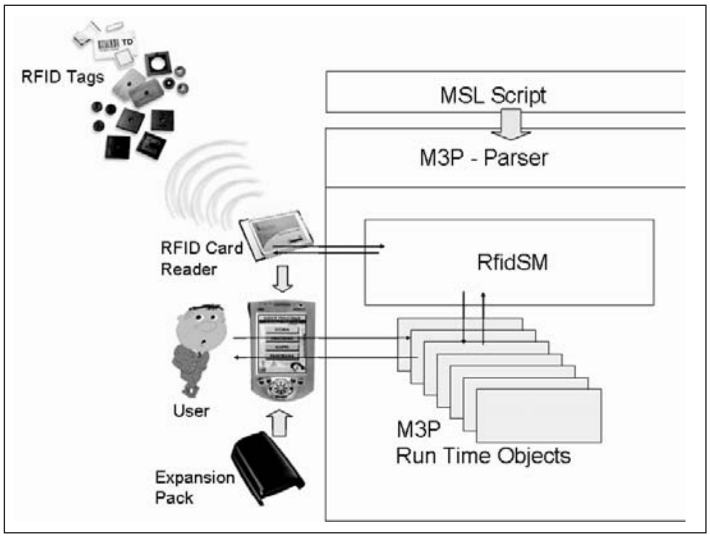
The script interface of the RfidSM is a new MSL component, called RFID, that exposes the fields shown in Table 1.

When the RfidSM component is started, and until it receives a stop event, it scans the surrounding environment to check the presence of tags every "period" of time. The list of detected tags is then sent with the MADE message-ex-

Table 1. RfidSM fields description

Component Field	Description
Target	List of identifiers of the components to which information about identified tags are sent
Period	A time period in milliseconds between two consecutive environmental scans to detect tags
Repetition	A number of tag detection operations executed consecutively on each scanning action
Id	A list of RFID tags that are of interest for the component
Delay	A list of time frames, one for each interesting tag, in which tags are not identified again
dBm	A list of signal strength values, one for each interesting tag, that specify thresholds for tag identifications
onFound	A list of events, one for each interesting tag, that RFID component launch when a tag is identified
Start	If a component launch this event on a RFID component starts the scanning of tags
Stop	If a component launch this event on a RFID component stops the scanning of tags

Figure 5. Integration (within MADE) of the RFID sensing module (RfidSM)



change modalities to the components specified in the "target" field. In addition, the component has an "id" field to allow programmer expressing interest in a set of tags, and defining (through the field "onFound") the corresponding events list that should be executed. Each interesting tag is also featured with a signal strength threshold (through the field "dBm") that specifies a limit under which the tag is considered in range.

There is the problem of collisions, since the scan results are typically imperfect due to not all tags are detected in every scan. To solve this problem, a tag typically awaits a random number of time slots before it answers the RF pulse sent by the reader. However, the problem still remains and grows as the number of tags in the surrounding environment grows. The MADE RFID sensing module tackles this issue, allowing the programmer to specify, through the field "repetition," a number of times that the reader should repeat the scanning before returning to the founded tags. The list of founded tags is the collection of all tags observed in each scan. A small value of repetition results in a fast scan with high risk of collision, whereas large repetition value results in a slow scan with few collisions. This trade-off should be resolved by the programmer basing his decision on application constrains: long delays can result in human-computer interaction problems if the application allows a user expectation for immediate reaction to tags. This is the case of applications in which user a voluntarily accosts the mobile device to tagged objects to obtain information. Similar problems arise if the application has a short time frame to detect tags, for example, in applications where the user moves at relatively high speed in the environment, like in territorial games. Instead, others type of applications can gain advantage from precise but slow detections. It is the case of a tourist mobile guide for a naturalistic park in which the user moves along a path with some points of interest largely spaced, like tree species or rare flowers.

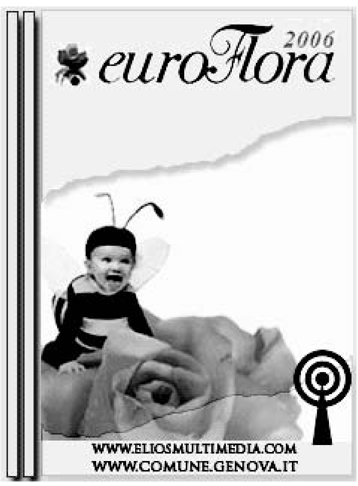
The other problem affecting the RFID technology is the "tag detection flickering" (Römer, Schoch, & Mattern, & Dubendorfer, 2003): due to the collision problem, some tags can appear and disappear from sequential scanning, generating a fast list of tag identifications. The MADE RFID sensing module allows the programmer to decide how to convert scan results into applications events handling this problem. Programmer can specify, through the "delay" field, a time period (for each interesting tag) starting from the detection. During this time, subsequent detection events of the same tag are discarded; also, the exact definition of this delay is application dependent. Applications with events that occur only one time, like tourist guide for museums with linear path, can have delay values set to infinite. Instead, in applications with events generated multiple times closer each other, like territorial games, the delay should be short or zero.

Currently, we have implemented the low-level driver support for the iCARD Identec reader in a PCMCIA card format (IDENTEC). This card can be integrated in handheld, portable, or laptop computers to communicate with the iQ and iD active RFID tags at a distance of up to 100 meters. The RF signal is in the UHF radio band (915 MHz or 868 MHz), providing long-range communication and high-speed transmission rates for reliable data exchange.

THE EUROFLORA GUIDE

In order to assess the possibility of using RFID technology to develop widely usable interfaces, we present a real-world application (see Figure 6) developed through the MADE toolkit and deployed in an ecological environment (fully operational, reliable, and robust application, used by real users and in a real context of use) at EuroFlora 2006 (the international flower exhibition that is held in Genoa every 5 years). With over 500,000 visitors in 10 days, EuroFlora is one of the most important exhibitions of Europe.

Figure 6. Snapshot of the cover page of EuroFlora Guide application



The developed application concerns the research area of assistive technologies for visually impaired people. Such assistive applications have the potential to improve the quality of life of a large portion of population (by 2020, there will be approximately 54 million of blind persons over age 60 worldwide (WHO, 1997)).

In this field, maintaining spatial orientation is a major challenge for people with visual impairment. There is the need of systems in providing blind people with information on where they are, hazards that might be in the way, and a description of what lies in their surroundings (Mann, 2004). The notion of "spatial orientation" refers to the ability to establish awareness of space position relative to landmarks in the surrounding environment (Guth & Rieser, 1997). The goal of our application is to support functional independence to visually impaired people, providing support to indoor awareness of elements in the surroundings (Ross, 2004).

The EuroFlora guide is organized in two parts. One part provides general information about the exhibition, the guide, and their services. The other part provides the description of the selected interest points. While first part is directly accessible by the user at any moment, the second one is event driven. More precisely, every interest point description is associated to an RFID tag, and when a user enters that area (i.e., her/his handheld device recognizes the RFID tag), the software asks the user whether to launch the corresponding description.

We placed 99 RFID sensors on an area of 30,000 mq of exhibition, covering 99 points of interest, services, and major areas (see Figure 7). RFID sensors were IP65 compliant in order to resist to water and dust, and self-powered. Power level of sensors could be set in two levels, low and high.

Design Methodology

The necessity for combining the flexibility and multimedia potential of a mobile device with the

Figure 7. a) The packaging of the multimedia guide in a leather case; b) Snapshots from the tests: users visit EuroFlora 2006 supported by the guide and touch some dedicated plants

extreme simplicity of interaction, required for use by a wide audience (also visually impaired people), involves facing three main HCI issues:

- **Usability by general users:** The tourist has little time and willingness to learn how to use the new technological tool, since she or he is there to visit the exhibition and not to learn a tool. Most of the tourists use such a tool for the first time and just for a short time (typically, from 30 to 90 minutes). Thus, the complexity of the platform should be hidden from visitors, making the guide immediately usable, with no effort by users. This implies that the interface is to be as simple and intuitive as possible.
- **Usability by visually impaired people:** Visiting an exhibition is a critical task for the blind, mainly for the combination of several reasons: the site is often crowded and unfamiliar to the visitor, it may be noisy, it is difficult to orientate in a highly dynamic place. In this context, the guide should be not intrusive, with few and very recognizable input interface elements (also with tactile feedback), and should give information in a proactive modality when needed by the user.
- **Presentation of information:** Added-value information (e.g., how the various specimens live in their natural environment) should be synergistic with the direct experience of the visitor at the exhibition. Provision of information has to be structured in order to enhance the direct perception of the visitor, leading to a better and more pleasant comprehension of her/his surrounding environment. For example, the guide should make use of environmental sound (e.g., waterfall) and scent (e.g., flower smell) to connect content information and the objects in the space.

We have tackled such issues resorting to the methodologies of the user-centric design (Carroll, 1997), in an iterative development of the guide involving participatory design (Beck, 1993), definition of usability specifications, and contextual design, as shown in the following:

- Participatory design consisted of the participation of botanists at the design decisions, authors skilled in writing for blind people and visually impaired end-users, together with technical developers. The most significant contribution of the first three categories consisted in the definition of the targets and in the concrete perspective they brought into the project.
- Usability specifications provide explicit and measurable targets to verify the suitability of the work done. Examples of such goals are "90% of the users should be able to operate the guide without asking questions to the personnel," "90% of the users should be able to use the interface with no errors," "90% of the users should be able to understand the

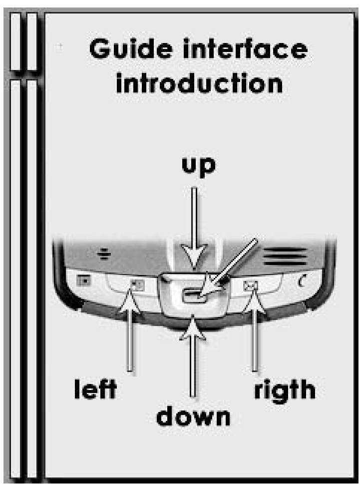
meaning of all the touchable controls within 120 seconds." All these objectives were verified in early lab and field tests in order to take the appropriate corrective actions.
- Contextual design involved early field tests with experts and users at the exhibition in the preopening days when the infrastructure of EuroFlora was being built. Field tests have been helpful to highlight problems and shortcomings that had been overlooked or ignored in lab.

Structure of the Interface

The interface of the EuroFlora guide has been designed to support immediate use by the general public, also by visually impaired people. To this end, we used general design principles (we already described them in the introduction) such as overall simplicity, low intrusiveness, and support for natural interaction and knowledge acquisition. Moreover, we added further features in order to meet the specific needs of visually impaired people:

- Tactile feedback in the control interface
- Tutorial stage
- Event-driven interface
- Facilities to support orientation

Figure 8. Snapshot of the guide instruction



The basic element of the interface is the multimedia card. A multimedia card corresponds to each subject of a presentation (e.g., a flower species). Each multimedia card provides, in an audio format, texts specifically written for visually impaired people (i.e., highlighting olfactive and tactile sensorial information, providing detailed ambient descriptions).

The tactile feedback is necessary to allow impaired people to easily understand the position of the controls and give her/him feedback. Our previous multimedia guides had the interface embedded in the graphic contents, exploiting the touch screen of a pocket-pc device. During the early field tests, visually impaired people pointed out some important shortcomings in these solutions. They felt that the screen was too large and their fingers were lost in a space without roughness. Since most of such users are well acquainted with the common cell phones' relief keyboard, we studied a new solution exclusively based on the hardware buttons (Figure 8).

The hardware buttons of the interface are highlighted. The choice of this area as navigation control allows visually impaired people to have a tactile feedback. The meaning of the button is "up" to accept a content description (which is automatically triggered when the user enters a new cell), "down" to reject it, "right" to exit from a section of the guide, and back to the main menu, "left" to have contextual information about user's current position.

The tutorial stage is an initial guide section in which users could freely experiment withthe interface of the tool in order to allow people to use the guide in an independent way. In this stage, users are invited to freely press buttons. A speech description briefly explains the meaning of each pressed button. This tutorial stage prevents the necessity for providing papers or long explanations when users rent the guide.

The event-driven interface allows a user to get information about points of interest (POIs) and orientation when they are in the proximity of a POI. For example, in Figure 9, a user near the azalea stand is told about the presence of this flower by a pop-up window (the guide is to be

Figure 9. Snapshot of the event-driven interface

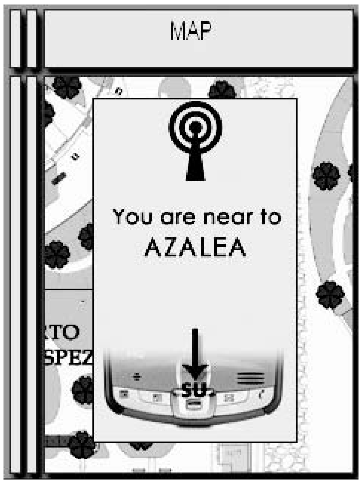
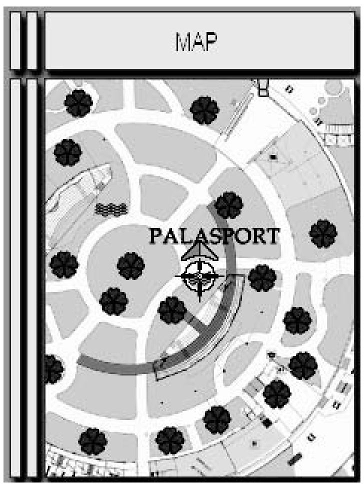


Figure 10. Snapshot of the map section of EuroFlora Guide



usable by everybody) and a corresponding jingle sound. If she/he wants to listen to this content, she/he can press the "up" hardware button. By default, the system skips the presentation. This operational mode has low intrusiveness (users are asked whether to listen to a content), but it also provides a certain degree of proactivity. Information is not only botanical, as in the example of azalea, but also concerns the positioning of the user. Many tags are placed in the proximity of facilities, such as lavatories, cafés, exits, and intersections. This localization system lowers the mental workload necessary for the tourist to synchronize the physical space of the exhibition with the virtual space of the program.

This message (accompanied by a jingle) is shown to the user when she/he has just entered a POI area. The combination of audio and graphics is due to the fact that the guide may be used also by not visually impaired people. In the example in this figure, the user is near to the azalea flower, and if she/he is interested in the description she/he can press the "up" hardware button to access the related content.

One of the main tasks of the guide is to assist the visitor in her/his exploration of the exhibition space. A facility to support orientation (not useful for visually impaired people) is a section with the map that helps the tourist to orient herself /himself in the exhibition. The map (see Figure 10) shows the structure of the EuroFlora, including lavatories, café, exits, and so forth, and the location of the points of interests. In order to enhance user's orientation, the map is centered on the position of the user, as determined by the currently perceived RFID tags.

FIELD EVALUATION

Experimental Framework

Real evaluation of advanced mobile device applications and of the impact on their intended population is difficult and costly. Evaluation requires analysis of real users, in a real context of use. In order to adequately evaluate interaction with computation resources, test-people should use a fully operational, reliable, and robust tool, not just a demonstration prototype (Abowd & Mynatt, 2000). Hence, it is important to perform early tests in the authentic context of use in order to verify end-user acceptance and overall usefulness of the system, and to receive feedback to inform future design.

In this chapter, we describe the early stage analysis of acceptance and usefulness of the developed multimedia guide. The tests were performed at EuroFlora 2006, the international flower exhibition that is held in Genoa every 5 years. With over 500,000 visitors in 10 days, EuroFlora is one of the most important exhibitions of Europe.

The exhibition area (around 90,000 squared metres) was equipped with an infrastructure of 99 RFID tags. The experimentation involved 120 visually impaired tourists who used the tour guide, and were observed and interviewed by expert of disabilities and HCI designers. Subjects were visually impaired (25%) or blind persons (75%) representing a wide range of age (28% age <30; 32% age between 30 and 50; and age >50 40%). Moreover, the tests involve 64 females and 56 males.

The tour guide consisted of a PocketPC equipped with a special leather package with a lace dropping from the neck for a more comfortable use. Headphones were also used in order to isolate the user from the highly noisy surrounding environment (see Figure 1).

Preexhibition Tests

In an early test session-performed 2 days before the official opening of the exhibition, when some stands were already readyenabling a realistic test-we prepared a prototype software version that was used by five selected visually impaired users visiting 30% of the total exhibition area. We followed and interviewed the users in this phase, in order to understand shortcomings, defects and weaknesses, and strong points of the product. In this phase, we understood and solved some problems on user interface and contents, such as the most suited assignment of buttons to presentation control functionalities and the length of the descriptions. Some test-users found the long silence time between a presentation activation and the next one (i.e., the period of time in which the user is walking through areas not covered by RFID tags) frustrating. We partially tackled this issue by periodically providing a message saying that the user is currently in an area not close to a POI.

Ecological Tests

One hundred and twenty blind people used the guide during the exhibition. Sixty of them (aged from 12 to 78 years old) participated in an ecological test conducted by the authors. We interviewed the users at the return of the guide. We evaluated three main performance factors: usability (including effectiveness, efficiency and pleasantness of use), usefulness, and capability to support spatial orientation (in particular the approach to the POIs). We asked users to give a general comment on the guide and a 1-5 grade for each factor (which was carefully explained and defined by the interviewers). An overall survey of results is reported in Table 2; it clearly shows the high acceptance by the users.

Analyzing the variables' correlations based on the chi-square test, we observed that usability is correlated with the perceived support for spatial orientation (χ=25.3, df (degree of freedom) = 16, 90% confidence), and that perceived utility of the tools is strictly correlated with perceived support for spatial orientation (χ=30.2, df=16, 99.9% confidence). This suggests the importance of our design choice to use mobile technology to support orientation of visually impaired people. Moreover, test results also show that the tool is perceived as useful and usable.

Considering the free comments, the guide was judged as an excellent tool for users to orientate themselves inside the exhibition. Several people

Table 2. Overall survey results

Issue	Average	Standard Deviation
Usability	4.00	0.64
Usefulness	4.25	0.75
Support for spatial orientation	4.20	0.66
Session length time	201 minutes	30 minutes

expressed a similar concept, which we can synthesize with the words of one visitor: "after always having been guided, for the first time I myself have been able to guide my wife and to explain the exhibition!" Such positive comments were also confirmed by the blind assistance experts, who highlighted the significant degree of independence the blind could reach through the guide.

Shortcomings in the interface were reported by some elderly users, while some people asked for more extended descriptions, though each point of interest included at least one. The high performance and reliability of hardware, software, and batteries assured long sessions of use with no troubles for the user.

FUTURE TRENDS AND VISION

The research community is envisaging a new model of a "tagged world" as an intelligent environment that allows providing visually impaired people with information about architectonical barriers, safe paths, points of interest, potential danger areas, and other useful information. A sample scenario description may give an idea of this likely future.

Maria is visually impaired. She is in a foreign city on a business trip. Maria owns a mobile device with a mobility-assistance system (MAS: it is similar to the EuroFlora Guide, but with a much larger action range). The MAS accompanies her in her path to her destination office, and signals pedestrian crossings, traffic lights, safe paths in work-in-progress areas, and so forth. All objects in the world send their signals, but Maria's wearable device has an intelligent reasoning algorithm (based on user preferences and interpretation of the user's current activity) and a suitable human-computer interaction (HCI) in order to provide her only with the needed information. This information is extracted from a mass of data that are continuously received from the close-by RFID tags. Thus, the wearable device notifies Maria about a pedestrian crossing only if it knows that this is useful for her current activity (i.e., going to office). Not useful information will not be provided, in order not to distract Maria. Along her path to her destination, Maria passes by a newsagent. The World Guide scans all the magazines and identifies today's issue of Maria's favourite magazine. It queries Maria's database, which replies that Maria has not purchased this issue yet; so, it notifies her about the opportunity to buy the magazine.

CONCLUSION

The ubiquitous presence of smart tags will offer, in the near future, a critical mass of information, embedded in the world, that will be exploitable to rethink the relationships between people involved in their daily-life activities and the surrounding world.

With MADE we have designed a system that continuously scans the tagged world, interprets the large amount of information coming from the surrounding objects, and provides it to the users through multimedia human-computer interaction. Moreover, the application in the future will filter the raw data coming from the environment (with artificial intelligence behaviour) taking into account the user needs, preferences, and profile.

The field test at EuroFlora 2006 has demonstrated the feasibility of our vision, by deploying the system in a real-world setting (an exhibition area with indoor and outdoor instrumented environments), and performing extensive field tests with real users. In a longer-term view, with such an application, we intend to investigate the future scenarios that will be enabled by a massive presence of RFID tags in our environments. This "early prototyping" has allowed us to understand, as early as possible, costs, limits, strengths, and benefits of the new technology. We have also obtained a significant positive feedback on user acceptance. Usability results show that the guide is perceived as highly usable and useful, in particular because of its ability to support spatial orientation.

The next step towards a "tagged world" will require integration of data and services, and capability of interpreting a variety of sources according to the specific and dynamic user needs. Achieving these goals will involve a huge research effort that will be successful only if it will lead to the deployment of compelling applications that will be perceived as useful by the users. In a user-centered design view, this implies a rapid prototyping of applications and extensive user testing in the real context of use, which was our inspiring principle in the EuroFlora project.

REFERENCES

Abowd, G. D., & Mynatt, E. D. (2000). Charting past, present, and future research in ubiquitous computing. *ACM Transaction in Computer-Human Interaction, 7*(1), 29-58.

Beck, A. (1993). User participation in system design: Results of a field study. In M. J. Smith, *Human-computer interaction: Applications and case studies* (pp. 534-539). Amsterdam: Elsevier.

Bellotti, F., Berta, R., De Gloria, A., Ferretti, E., & Margarone, M. (2003). VeGame: Field exploration of art and history in Venice. *IEEE Computer, 26*(9), 48-55.

Bellotti, F., Berta, R., De Gloria, A., Ferretti, E., & Margarone, M. (2004). Science game: Mobile gaming in a scientific exhibition. eChallenges e2004. *Fourteenth International Conference on eBusiness, eGovernment, eWork, eEurope 2005 and ICT.* Vienna.

Bellotti, F., Berta, R., De Gloria, A., & Margarone, M. (2002). User testing a hypermedia tour guide. *IEEE Pervasive Computing*, *1*(2), 33-41.

Bellotti, F., Berta, R., De Gloria, A., & Margarone, M. (2003). MADE: Developing edutainment applications on mobile computers. *Computer and Graphics*, *27*(4), 617-634.

Borriello, G., Chalmers, M., La Marca, A., & Nixon, P. (2005). Delivering real-world ubiquitous location systems. *Communications of the ACM, Special Issue on The Disappearing Computer, 48*(3) 36–41.

Bruno, R., & Delmastro F., (2003). Design and analysis of a bluetooth-based indoor localization system. *Personal Wireless Communications, 2775*, 711-725.

Carroll, J. M. (1997). Human-computer interaction: Psychology as a science of design. *International Journal of Human-Computer Studies*, *46*(4) 501-522.

Djuknic, G. M., & Richton, R.E. (2001). Geolocation and Assisted GPS. *IEEE Computer, 34*(2), 123-125.

EuroFlora. (s.d.). EuroFlora 2006 Home Page. Retrieved from http://www.fiera.ge.it/euroflora2006/index_eng.asp

Fleck, M., Frid, M., Kindberg, T., O'Brian-Strain, E., Rajani, R., & Spasojevic, M. (2002). From Informing to Remembering: Ubiquitous Systems in Interactive Museums. *IEEE Pervasive Computing*, *1*(2), 11-19.

Goker, A., Watt, S., Myrhaug, H. I., Whitehead, N., Yakici, M., & Bierig, R. (2004). An ambient, personalised, and context-sensitive information system for mobile users. *Second European Union Symposium on Ambient Intelligence*. Eindhoven, The Netherlands.

Guth, D. A., & Rieser, J. J. (1997). Perception and the control of locomotion by blind and visually impaired pedestrians. In *Foundation of orientation and mobility* (2nd ed.) (pp. 9-38). AFB Press.

Howard, A., Siddiqi, S., & Sukhatme, G. (2003). *An experimental study of localization using wireless Ethernet.* International Conference on Field and Service Robotics

IDENTEC. IDENTEC Solution Ltd Home Page. Retrieved from http://www.identec.com

Kindberg, T., & Barton, J. (2001). A Web-based nomadic computing system. *Computer Networks, 35*(4), 443-456.

Langheinrich, M., Mattern, F., Romer, K., & Vogt, H. (2000). *First steps towards an event-based infrastructure for smart things.* PACT 2000. Philadelphia.

Liu, X., Corner, M.D., & Shenoy, P. (2006, September). *RFID* Localization for pervasive multimedia. In *Procedings of the 8th International Conference on Ubiquitous Computing (UbiComp '06).* California, USA.

López de Ipiña, D., Mendonça, P., & Hopper A. (2002). TRIP: A low-cost vision-based location system for ubiquitous computing. *Personal and ubiquitous computing, 6*(3), 206-219

Mann, W. C. (2004). The aging population and its needs. *IEEE Pervasive Computing, 3*(2), 12-14.

Quaadgras, A. (2005). Who joins the platform? The case of the RFID business ecosystem. *38th Hawaii International Conference on Systems Science* (pp. 855-864). Gig Island (HI): IEEE Computing Society Press.

Römer, K., Schoch, T., Mattern, F., & Dubendorfer, C. (2003). T-smart identification frameworked for ubiquitous computing applications. *PerCom 2003.* Fort Worth: IEEE Press.

Ross, D. A. (2004). Cyber crumbs for successful aging with vision loss. *IEEE Pervasive Computing, 3*(2), 30-35.

Siegemund, F., & Floerkemeier, C. (2003). Interaction in pervasive computing settings using Bluetooth-enabled active tags and passive RFID technology together with mobile phones. *PerCom 2003.* Fort Worth: IEEE Press

Want, R. (2004). Enabling ubiquitous sensing with RFID. *IEEE Computer,* 84-86.

Want, R., Fishkin, K., Gujar, A., & Harrison, B. (May 1999). Bridging physical and virtual worlds with electronic tags. *ACM Conference on Human Factors in Computing Systems (CHI 99).* Pittsburgh, PA.

Want, R., Hopper, A., Falcão, V. & Gibbons, J. (1992). The active badge location system. *ACM Transactions on Information Systems, 10*(1), 91-102.

WHO, W. H. (1997). *Who sounds the alarm: Visual disability to double by 2020.* Retrieved from http://www.who.int/archives/inf-pr-1997/en/pr97-15.html

Xu, Z., & Jacobsen, H. A., (2005). *A framework for location information processing.* 6th International Conference on Mobile Data Management (MDM'05). Ayia Napa, Cyprus.

KEY TERMS

Chi-Square Test: The Chi-square is a test of statistical significance for bivariate tabular analysis (crossbreaks). This test provides the degree of confidence we can have in accepting or rejecting a hypothesis.

Ecological Context: The ecological context is a set of conditions for a user test experiment that gives it a degree of validity. An experiment with real users to possess ecological validity must use methods, materials, and settings that approximate the real-life situation that is under study.

Human-Computer Interaction: Human–computer interaction (HCI), also called man-machine interaction (MMI) or computer–human interaction (CHI), is the research field that is focused on the interaction modalities between users and computers (interface). It is a multidisciplinary subject, relating to computer science and psychology.

Location-Aware Computing: Location-aware computing is a technology that uses the location of people and objects to derive contextual information with which to enhance the application behaviour. There are two ways to acquire information about user context: requiring the user to specify it or by monitoring users and computer activity. Sensor technology, such as RFID, could enable mobile devices to extract information from user position automatically.

Mobile Tourist Guide: A mobile tourist guide is a software application with an intuitive interface, that provides users with multimedia information when and where needed during their visit to museums, city centres, parks, and so forth.

Such an application runs on PDA-type terminals or on cellular phones, and could be augmented with GPRS (general packet radio service), GPS (global positioning system), and Bluetooth wireless technology. The guide allows tourists to plan routes according to preferences and ambient conditions (weather, timetables, sites of special interest, etc).

Radio Frequency Identification: Radio frequency identification (RFID) is an automatic identification method based on storing and remotely retrieving data using small and cheap devices called RFID tags or transponders. An RFID tag is an object that can be attached to objects, products, or persons to identification using radio waves. Passive tags (with a few centimeter range of sensitivity) require no internal power source, whereas active tags (with more long range of sensitivity, 100 meters) require a power source.

User-Centric Design: User-centric design is a design process that aims at realizing products that meet users' expectations. The key idea of this design methodology is to start the design strategy taking into account the user's perspective.

Chapter XL
Toward a Novel Human Interface for Conceptualizing Spatial Information in Non-Speech Audio

Shigueo Nomura
Kyoto University, Japan

Takayuki Shiose
Kyoto University, Japan

Hiroshi Kawakami
Kyoto University, Japan

Osamu Katai
Kyoto University, Japan

Keiji Yamanaka
Federal University of Uberlândia, Brazil

ABSTRACT

We developed a concept of interfaces using nonspeech audio for building wearable devices to support visually impaired persons. The main purpose is to enable visually impaired persons to freely conceptualize spatial information by nonspeech audio without requiring conventional means, such as artificial pattern recognition and voice synthesizer systems. Subjects participated in experiments to evaluate their ability to localize pattern-associated sounds. During the experiments, the subjects navigated through various virtual 3-D acoustic environments. The experimental results showed that sound effects, such as reverberation and reflection and variable z-coordinate movement, enhance the ability to localize *pattern-associated sounds. The subjects were also evaluated on their ability to conceptualize spatial information based on cues in "artificial" and "natural" sounds. The evaluation results revealed that "natural" sounds are essential for improving everyday listening skills and the ability to conceptualize spatial information.*

INTRODUCTION

Novel concept of interfaces are needed to keep pace because computing devices drop in size and rise in power/bandwidth. Interfaces that rely on screens and keyboards are not always effective in mobile computing scenarios. Screens and keyboard-based interfaces are difficult to use whilst the arms and hands are involved in real-world tasks (Brewster & Walker, 2000). Brewster and Walker (2000) looked at how current interfaces to handheld computers can be improved by the use of audio, and developed purely auditory interfaces with gestural control to avoid the need for visual attention.

According to Gibson (1966), a system has organs, whereas a sense has receptors. A system can orient, explore, investigate, adjust, optimize, resonate, extract, and achieve an equilibrium. The term "sense" means "to detect something," which is more accurate than "to have a sensation" in senses considered as perceptual systems. While the achievements of a perceptual system are susceptible to maturation and learning, special sense inputs constitute a repertory of innate sensations. Sensations can be organized, fused, supplemented, or selected, but they cannot be learned. In a perceptual system, the input-output loop can be assumed to actively obtain information that becomes more and more subtle, elaborate, and precise with practice. One can keep on learning to perceive as long as life goes on.

Gibson (1966) also conjectured that perceptual systems develop perceptual skills analogous to the way in which behavioral systems develop performative skills. Perceptual systems are amenable to learning, but the channels of sense are not subject to modification by learning.

We assume sensations triggered by sound are merely incidental, sound information is available to perceptual systems, and the qualities of the real world in relation to the needs of subjects are experienced directly.

On the basis of these assumptions about perceptions and senses, we analyzed the learning processes of subjects, considered as perceptual systems, during sound localization and spatial conceptualization experiments.

For most animal species, accurate sound localization is ecologically important since it is fundamental for survival, communication, and learning about sight-sound correspondences (Fujiki, Riederer, Jousmaki, Makela, & Hari, 2002).

The ability to localize a sound in natural space is present in nearly all animals that possess a hearing mechanism (Erulkar, 1972). However, despite the practical significance of this ability, our knowledge about the development of sound localization skills in humans is fairly limited. Sound localization ability has been exploited in our work as a background to reach a sound visualization process by subjects.

Plenge (1974) has investigated that "localization" refers to judgments of the direction and distance of a sound source. The sound image can be located inside the head by wearing the headphones. On the other hand, "lateralization" describes the apparent location of the sound source within the head. Headphones play important roles, allowing the precise control of interaural differences to eliminate sound effects related to room echoes. Therefore, lateralization may be regarded as a laboratory version of localization that provides an efficient means of studying sound direction perception.

In the conceptualization process, researchers have shown that sounds more complex than ubiquitous interrupting beeps are necessary to provide spatial information to computer users. The evaluated vOICe Learning Edition (Jones, 2004) is an actual example of interface that translates

Figure 1. Conventional human-environment interface

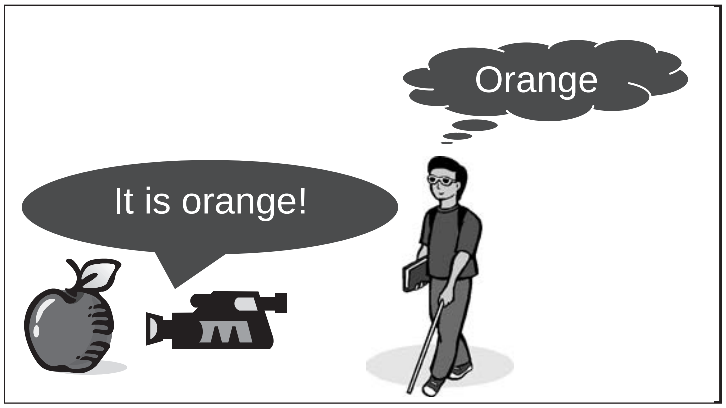
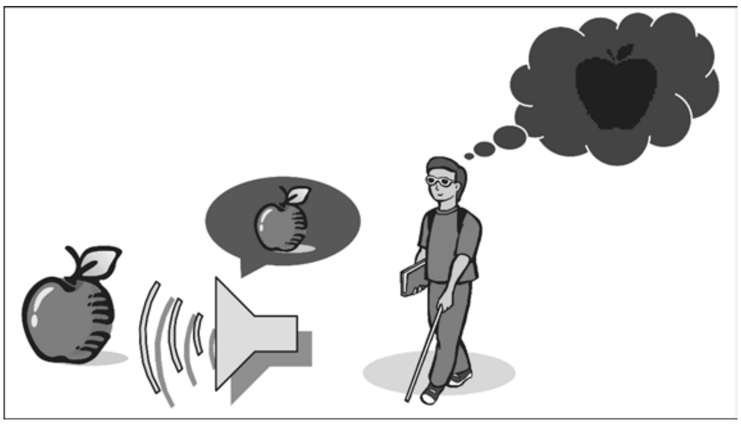


Figure 2. Novel human-environment interface



arbitrary video images produced by an ordinary camera into sounds. However, like vOICe, the problem is that the sounds similar to ubiquitous interrupting beeps have been used to represent the data characteristics. This requires extensive trials before actually using the interface because the sounds have no analogs in everyday listening.

According to Buxton (1990), we have highly developed "everyday listening" skills to everyday tasks such as driving and crossing the street. In other words, everyday listening (Buxton, 1990; Gaver, 1993) is the experience of listening to events rather than sounds. Listening to events is hearing which things are important to avoid and which might offer possibilities for action. This experience is different from listening to music (perceptual dimensions of the sound itself), and is unfortunately not well understood by traditional audition approaches.

Improving everyday listening skills has virtually been neglected at the expense of interacting with computers. We propose a comprehensive account of everyday listening by analyzing different sounds to avoid this waste of the audio channel and improve the spatial conceptualization of interface users.

WORK'S NOVELTY

We break away from the conventional use of speech audio and "artificial" sounds toward developing novel interfaces based on nonspeech audio and "natural" sounds that are heard everyday.

Our concept is based on the skill transfer process (Shiose, Sawaragi, Nakajima, & Ishihara, 2004) for building paths to pattern-associated sound localization and spatial conceptualization. We aim to solve the problems of the conventional concept by providing users, such as visually impaired persons, with the opportunity to perceive and embody sound cues without need for "expert systems" that require heavy computational load.

Conventional human-environment interfaces, as shown in Figure 1, primarily consist of such expert systems:

1. An image preprocessor for scanned images
2. An image segmentation system for obtaining patterns
3. An accurate pattern classification system for categorizing different patterns
4. A converter of pattern information into spoken language
5. A voice synthesizer system

Our novel human-environment interfaces, on the other hand, would provide users with the freedom to conceptualize the target (pattern or environment) and interact with it using valuable human perceptual skills. These interfaces would be free of expert systems, such as accurate image classification systems, spoken language converters, and voice synthesizers, which require a heavy computational load that is not supported by current mobile technology. In other words, users would interact with the target by freely conceptualizing spatial information as shown in Figure 2.

We first evaluated subjects' abilities to localize sound. We generated pattern-associated sounds with effects, such as movement, reverberation, and reflection. The sounds were based on the real-world sounds and were simulated with a virtual 3-D acoustic space system. In other experiments, we evaluated subjects' abilities to conceptualize spatial information as they navigated through a virtual 3-D acoustic space while listening to "artificial" and "natural" sounds.

During these experiments, we evaluated subjects' abilities to perform tasks for localizing pattern-associated sounds and conceptualizing spatial information while listening to nonspeech sounds. We believe interface users can embody the same skills, used in such everyday tasks as crossing the street, by navigating through a virtual 3-D acoustic space while listening to nonspeech sounds to conceptualize spatial information. Our "built-up association" of reverberations with empty space is an example of such skills. The more sound reverberance in a room, the more empty space it has, all other things equal. This built-up association should provide a natural way of conceptualizing spatial information based on experience and familiarity with everyday listening. Our sound system for creating a virtual 3-D acoustic space enables interface users with highly developed everyday listening skills to perform everyday tasks without requiring sophisticated devices.

Our concept enables the development of user-friendly devices that do not require hard training and a heavy computational load. Such devices would be useful to blind people for conceptualizing spatial information as glasses are to people who require vision correction.

ADVANTAGES OF USING OUR EARS

It is worthwhile to compare the relative speeds with which the ear and eye attain their maximum sensitivity. According to Handel (1989), the eye takes more than 30 minutes in darkness to attain its maximum sensitivity while the ear reaches its maximum in only 0.1 second.

Kallinen (2003) has investigated new ideas for using audio for feedback and for presenting and managing information in computers from the perspective of the structure of sounds and the semiotic theory of signs. The motivation is that the auditory modality, such as speech, signals and natural sounds, is one of the most important ways to present and communicate information, but in conventional computer interfaces, these ways have been almost totally neglected.

From a neuroscience view, Corey and Hudspeth (1979) have investigated the sensitivity of the ear, concluding that the bundle of stereocilia works as a light switch. The bundle works by turning the cell on when it moves from the shortest cilia to the tallest one. But when the bundle moves in the opposite direction, it turns the cell off. Hair cells are so sensitive that deflecting the top of a bundle by the width of an atom is enough to make the cell respond. This deflection is equivalent to displacing the top of the Eiffel Tower by only half an inch, which might be caused by a very low, quiet sound at the threshold of hearing.

Another component of sensitivity is our ability to distinguish changes in sound. Experiments using pure sounds have measured changes in intensity or in frequency required before subjects can judge two sounds as being different. Investigators have concluded that the hair cells' response is amazingly rapid. For instance, we can accurately discriminate between highly similar sounds even from birth: the slight difference between the sounds "pah" and "bah" is noticed by newborn infants (Eimas, 1975).

From Kendall (1991), the combination of visual and auditory imagery offers a way of presenting and communicating complex events that emulates the richness of daily experience. In everyday life, sound events arise from action, that is, from the transfer of energy to a sound object. Even in childhood, we learn to recognize the occurrence of sound events and to relate them to physical events. We learn to classify and identify heterogeneous sound events through a lifetime of experience. Simple categorical distinctions of sound events can be exploited in auditory presentations to communicate important distinctions in the data. Scientific visualization (Kendall, 1991) is an example of such auditory presentation for communicating data. The primary goal of this scientific visualization is to provide scientists with improved representations of complex phenomena, and recent advances in multimedia and virtual reality have opened up other ways to interpret data.

Sounds play a more important role in the study of these complex phenomena through the use of auditory data representation and "visualization" by ear (Buxton, 1990; Kendall, 1991).

According to Yeung (1980), such human sensations as taste, smell, heat, and touch are not suitable channels for data presentation because our perception of them is not quantitative. However, sound is a useful medium for data presentation. Yeung (1980) explored the use of sound as an alternative to the graphic presentation of data because there are several problems associated with visual presentations as follows: adequate standards are generally not available for visual displays; resolution of the representation is generally poor; there are problems with the actual orthogonality of the axes in visual representations; the problem remains of scaling the measurements in each dimension.

A comprehensive account of everyday listening has yet to emerge. It is necessary to develop an account of ecologically relevant perceptual entities, and the dimensions and features of events that we actually obtain through listening: that is, "What do we hear (Gaver, 1993)?" Other listening-based studies (Gaver, 1993; Heine & Guski, 1991; Shaw, McGowan, & Turvey, 1991; Vanderveer, 1979; Warren & Verbrugge, 1984), from an ecological perspective, suggest that an ecological approach to audition could be fruitful.

Our previous works (Nomura, Yamanaka, Katai, Kawakami, & Shiose, 2004; Shiose, Ito, & Mamada, 2004) concentrated on such everyday listening. Our work, which concerned hearing an approaching automobile, is useful in understanding what we hear and thus the scope of an ecological approach to audition (Shiose et al., 2004). Experimentally, a source event (automobile) caused sound waves, some of which radiated directly to an observation point and others were modified by the environment before being reflected onto the subject. The experience of everyday listening captured by the auditory system is based on the idea that a given sound provides information about an interaction of materials at a location in an environment. In an analogous fashion to changes in visual texture by Shaw et al. (1991), we believe that changes in loudness caused by changes in distance from a source may also provide information about time-to-contact. A goal is a system for training the ability to pick up information about this time-to-contact to support visually impaired persons.

ECOLOGICAL PSYCHOLOGY APPROACH TO AUDITORY EVENT PERCEPTION

According to Gibson (1979), ecological psychology studies human-environment interrelationships and human perception in rich environments. Perception is usually of complex events and entities in the everyday world. Sound provides information about the interaction of materials at a place in an environment. For instance, we can hear an approaching automobile, its size, and its speed. These are the phenomena concerning an ecological approach to perception. This perception is direct and unmediated by inference or memory. Elemental stimuli for perception do not necessarily correspond to primitive physical dimensions, but may instead be specified by complex invariants of supposedly primitive features. Complex perceptions rely on seemingly complex stimuli (or "perceptual information"), not on the integration of sensations. Since exploration of the world over time becomes an important component of perception, our descriptions are no longer limited to primitive physical dimensions. They should uncover the ecologically relevant dimensions of this perception and their invariant perceptual information from an ecological account.

Typical research has little to say about such information from an ecological account because it focuses on the sound itself, analyzing it in terms of such properties as amplitude and perceived loudness, or frequency and perceived pitch. Traditional researchers do not consider the higher-level structures, such as information about events.

To enable us to pick up such information, an ecological approach implies analyses of the mechanical physics of source events, the acoustics describing the propagation of sound through an environment, and the properties of the auditory system. These analyses consist of a characterization of acoustic information about sources, environments, and location which can be empirically verified.

BENEFITS OF NONSPEECH AUDIO

Synthetic speech audio is a very common means of delivering information to visually impaired users, but it can be laborious and time consuming to listen to and compare many values through speech alone. On the other hand, nonspeech audio is a better means at providing an overview of the data, as it can be delivered in a shorter time than synthetic speech (Wall & Brewster, 2006).

Bronstad, Lewis, and Slatin (2003) have investigated whether the use of nonspeech audio cues can reduce cognitive workload to the extent that users can perform very complex tasks that they would otherwise find impossible.

Our ears provide us a means to extract information from nonspeech audio that cannot be, or is not, displayed visually (Buxton, 1990). For example, knocking on objects tells us a great deal about the materials from which they were made. Nonspeech audio leads to important observations about the quality of materials compared to sight. Object-associated sounds emerge as the result of specific actions. There is latent information embedded in an object's sounds that can be exploited as a potent resource when transferred to computational objects. For example, the size of files, disks, and so forth is encoded in how high (small) or low (large) the resulting sound is in SonicFinder by Bill Gaver (1989). In another case, amount of reverberation associated with a sound provides cues about how full (dry) or empty (reverberant) the object (such as a disk) is.

According to Brewster, Wright, and Edwards (1993), the use of nonspeech audio at the user interface has become increasingly popular due to its potential benefits as follows:

- Information communicated to the user is increased.
- The amount of information received through the visual channel is reduced.
- Performance by sharing information across different sensory modalities is improved.
- Chances to identify the data are increased.
- Greater flexibility is provided because it can be heard from 360 degrees without needing to concentrate on an output device.
- A user's attention is captured better while he is performing another task.
- Humans are able to perceive changes in dimensions of sound in various degrees of effectiveness (Wise, 1993).

Blattner, Sumikawa, and Greenberg (1989) suggested synthetic timbres created by sine, triangular, and square waveforms. However, Brewster et al. (1993) emphasized that musical instrument timbres should be used instead of these synthetic timbres. In accordance with Brewster et al. (1993), we adopted a conventional violin to produce musical timbres as nonspeech audio source for the experiments in our sound localization process.

VIRTUAL 3-D ACOUSTIC SPACE

Handel (1989) has investigated the various perceptual consequences of the environment that all contribute to auditory perception: room size, reverberation time, reflection of sound waves, and source or listener movement. All of these effects modify the sound at the listener's ears from the sound source. We believe that sighted and nonsighted people can develop remarkable abilities to use reflected or reverberated sound patterns to move around and locate their position in rooms, corridors, and even outdoors, in spite of such irregularities. It is known that the sound-wave pattern changes in irregular ways even walking around a simple room. In other words, subjects can also locate moving and nonmoving sound-producing objects as well as nonsound-producing obstacles, and can often describe the size, shape, and texture of these objects from the pattern of the reflected and reverberated sounds. A sophisticated, learned skill demonstrates the potential of this information for understanding the external world.

The localization of sound depends on the way the sound waves from the same source differ from each other as they reach the left and right ear. The sound arriving at the ears can be modified by the head, torso, shoulders, and the outer ears. This modification can be described by a complex response function (head-related transfer function - HRTF) (Mφller, Sφrensen, Hammershφi, and Jensen, 1995). HRTFs contain all the information about the sound source's location (its direction and distance from the subject), and can be used to generate binaural cues (interaural time differences - ITDs; interaural intensity differences - IIDs) and monaural cues caused by the observer's own head and pinnae that spectrally color the sounds (Blauert, 1997). If properly measured and implemented, HRTFs can generate a virtual 3-D acoustic space.

A virtual 3-D acoustic space can be used to simulate a real-world environment; so it is a useful tool for investigating subject's abilities to localize sound and conceptualize their surroundings while they navigate through the virtual environment.

The system apparatus for creating the virtual 3-D acoustic space is only used for navigating and enabling the skill transfer process (Shiose et al., 2004) by eventual mobile device users. The sound generation system will not be implemented in the interface devices.

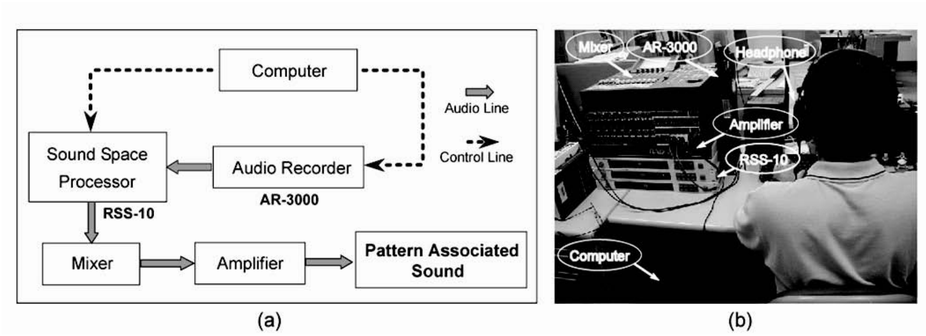
Our virtual 3-D acoustic space is based on a three-dimensional sound space processor (Roland RSS-10) that adds appropriate sound effects, such as reflection and reverberation, to a sound source to produce various nonspeech sounds for experiments.

SOUND LOCALIZATION PROCESS

Apparatus

A schematic overview of a virtual 3-D acoustic space system for generating pattern-associated

Figure 3. Sound generation system (Nomura et al., 2004): (a) a schematic overview; (b) front view photo



sounds is shown in Figure 3(a). This system is similar to our previous research (Shiose et al., 2004). As shown in Figure 3(b), each pattern-associated sound reaches the ears by headphone. Corresponding generated sounds depend on the "quality" (shape) of each pattern from an actual degraded image.

Experimental Process

In this work, the experimental process is quite similar to previous work (Nomura et al., 2004). Figure 4 shows a schematic overview of the experimental process to generate pattern-associated sounds with reverberation and reflection effects in the virtual 3-D acoustic space.

The process can be detailed as follows:

- Extraction of feature vectors from degraded images. The feature vectors as experimental data were obtained through an image pre-processing system (Nomura, Yamanaka, & Katai, 2002; Nomura, Yamanaka, Katai, & Kawakami, 2004) followed by character segmentation and feature vector extraction (Nomura, Yamanaka, Katai, Kawakami, & Shiose, 2005). The extracted feature vectors are 300-dimensional ones, and a sample of their shapes is shown in Figure 5.
- Dimensionality reduction of the feature vectors using multilayer perceptron (MLP) trained with the back-propagation algorithm (Haykin, 1999). A 3-D Cartesian coordinate system is used to locate the reduced feature vectors. Figure 5 presents the sample of 300-dimensional feature vectors and the corresponding reduced vectors (3-D vectors with x, y, z coordinates) (Nomura et al., 2004).

Figure 4. Schematic overview of the process for generating pattern-associated sounds

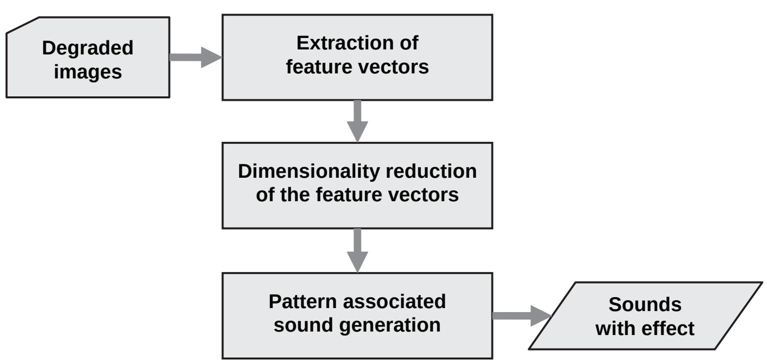


Figure 5. Sample of feature vectors and corresponding 3-D vectors for the experimental process (Nomura et al., 2004)

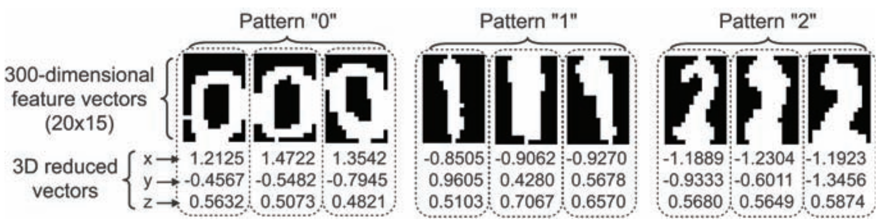


Table 1. Target vectors for training the MLP model

Vector	x-Coordinate	y-Coordinate	z-Coordinate
V1	1.5773	-0.4226	0.5773
V2	-0.4226	1.5773	0.5773
V3	-1.1547	-1.1547	0.5773

- Generation of pattern-associated sounds. Each sound source (produced by a conventional violin) was played by an AR-3000 audio recorder, and the corresponding sound effect such as movement, reverberation, or reflection was added using RSS-10 sound space processor as shown in Figure 3.

The Cartesian coordinates of the reduced vectors were multiplied by 10 to represent the distance in meters between the subject head and sound (position). Three-dimensional spatial geometry represented the movement direction of each pattern-associated sound to be perceived by users. The experimental conditions to the sound movement direction are as follows (Nomura et al., 2004):

- The initial position of the subject is (0, 0, 0) and he/she faces the 0*y* direction.
- The forward position of each sound is shifted 2 *m* ahead of the subject.
- The angle is calculated by a vector direction that links the initial position of each sound with the subject's head position.

Figure 6 shows the 2-D and 3-D movement directions for the training vectors presented in Table 1.

Experimentally, each subject works in two sessions. First, the subject is trained until he/she can capture the necessary skills to localize pattern-associated sounds. Second, the subject is tested to perceive the movement direction of several pattern-associated sounds and localize them. Details of the two sessions in each experiment are presented as follows by two data sets.

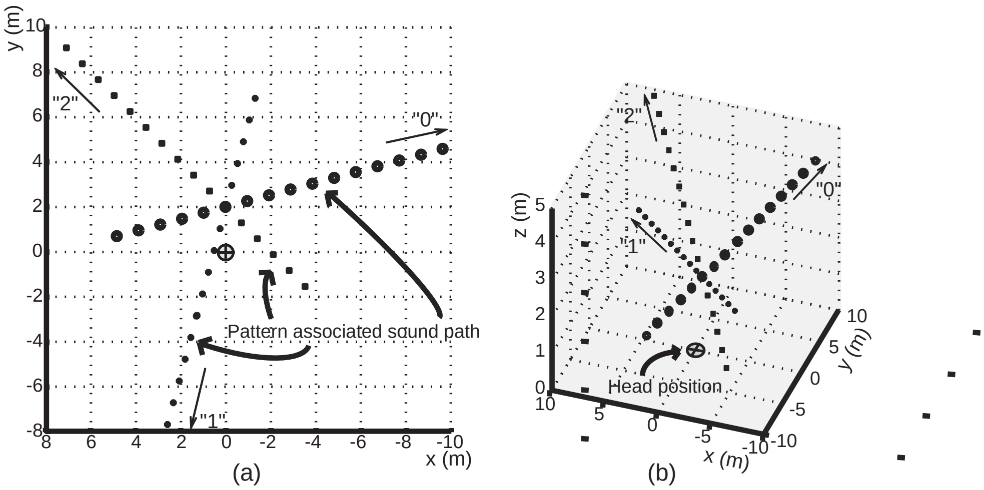
Training Session

Table 1 presents input data to train the subjects to acquire the necessary skills to localize respective pattern-associated sounds. But, the subject could

Table 2. Scale of musical notes (Nomura, Shiose, Kawakami, Katai, & Yamanaka, 2004)

G	G#	A	A#	B	C	C#	D	D#	E	F	F#	G	G#	A	A#	B	C	C#	D

Figure 6. Movement direction for pattern-associated sounds: (a) in 2-D coordinate system (fixed z-coordinate); (b) in 3-D coordinate system (variable z-coordinate) (Nomura, Shiose, Kawakami, Katai et al., 2004)



select the preferable musical note of sound source from the scale presented in Table 2.

Testing Session

Thirty 3-D feature vectors obtained by dimensionality reduction using MLP were used as input data in this session. The subject hears a pattern-associated sound with effects on sound source, and must associate it to symbol “0,” “1,” or “2.” In other words, the task for subject is to classify pattern-associated sounds by capturing spatial information cues in nonspeech sounds with effects.

Experimental Results

Subjects were requested to hear 30 testing sounds in virtual 3-D acoustic space. We obtained results to analyze the auditory event perception (sound localization) by these subjects. A group of 10 male and 3 female subjects ranging from 20 to 60 participated in the experiments. We randomly select 6 subjects and present their performance on sound localization task using various nonspeech sounds.

Experiment with the Fixed Z-Coordinate

A main condition for this experiment is pattern-associated sound moving with a fixed z-coordinate, that is, at a constant height like an aircraft cruising. Seven subjects participated in this experiment to provide sound localization results.

Experiment with the Variable Z-Coordinate

In this experiment, the sound movement condition was quite similar to an aircraft taking off from an airport, as shown in Figure 6 (b). This experiment is to verify the influence of sound height variation (variable z-coordinate) on the perception rate or sound localization ability of subjects. The subject’s task is to classify 30 pattern-associated sounds that move under variable z-coordinates.

Table 3 presents the detailed results corresponding to each sound localization performance of subjects in fixed and variable z-coordinate conditions.

Table 3. Detailed results to each pattern-associated sound as localization performance of subjects in experiments with the fixed and variable z-coordinates

		Pattern "0"			Pattern "1"			Pattern "2"		
Subject		3	4	5	3	4	5	3	4	5
Localization (%)	Fixed	70	10	20	30	20	30	30	60	10
	Variable	100	80	80	60	40	90	80	30	80

Experiment Including Reverberation Effect on Sound Source

Similar to previous work (Nomura et al., 2004), reverberation was attributed at a maximum level (0 *dB* attenuation), and reverberation time was defined as 0.2 *s*. Also, virtual room size was set to 4 *m* length for each wall.

Experiment Including Reflection Effect on Sound Source

Basically, reflection caused by the floor was set to a maximum level (0 *dB* attenuation) similar to previous work (Nomura et al., 2004) condition. Also, the height between the floor and subject's head was defined as 1.2 *m*. Results for sound reverberation and reflection influence on the localization rate by subjects 8, 9, and 11 can be verified in Table 4.

Discussion

The results listed in Table 3 show that the localization rate for the pattern-associated sounds moving with variable z-coordinates was better than that for those moving with fixed z-coordinates. For example, the data clearly shows a significant improvement in the localization rate (from 10 to 80%) for subjects 4 and 5 when they localized patterns "0" and "2."

In Table 4, the listed results show that subject 8 achieved the maximum localization rate (100% for all patterns) when both sound effects (reverberation and reflection) were included. After the test, subject 8 told us that he or she could readily hear the cues necessary to accurately localize the pattern-associated sounds produced with variable z-coordinates and both sound effects.

Better localization rates for pattern "0" than for patterns "1" and "2" suggest that subjects had difficulty discriminating the last moving sounds while the subjects navigated through the acoustic space. In other words, the results for localizing

Table 4. Detailed localization results by subjects in experiments with sound effects 1 and 2 on sound source, where effect 1 is "reverberation" and effect 2 is "reflection"

		Pattern "0"			Pattern "1"			Pattern "2"		
Subject		8	9	11	8	9	11	8	9	11
Localization (%)	With effect 1	100	100	100	90	60	80	90	50	30
	With effect 2	100	100	100	100	70	70	80	30	20
	With effects 1, 2	100	100	100	100	90	90	100	20	40

sound pattern "0" were better than those for localizing the other sound patterns because these patterns moved at the same side and were grouped together. This grouping enables subjects to reliably localize sounds in our 3-D acoustic space, but it can also be easily violated.

SPATIAL CONCEPTUALIZATION PROCESS

The process consists of the apparatus for creating virtual courses and experiments to investigate appropriate sounds for acoustic course (spatial information) conceptualization by subjects hearing "artificial" and "natural" sounds.

Apparatus

The apparatus is a virtual 3-D acoustic environment system based on sound space processors to generate the "artificial" and "natural" sounds for investigating the spatial conceptualization performance of subjects.

Figure 7 shows a front view and Figure 8 shows a schematic overview of the apparatus.

Figure 7. Front view of the apparatus for generating "artificial" and "natural" sounds (Nomura, Tsuchinaga, Nojima, Shiose, Kawakami, Katai, & Yamanaka, 2007)

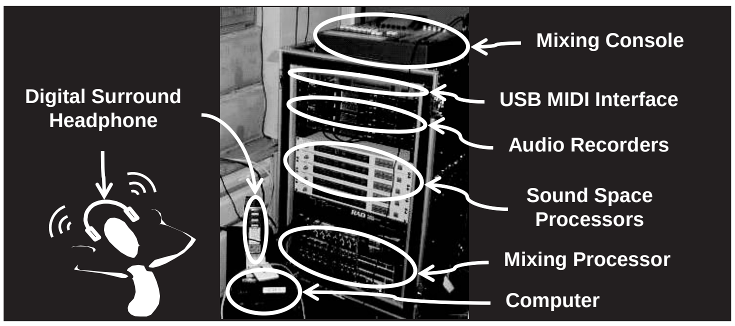


Figure 8. Schematic overview of the apparatus for experiments (Nomura et al., 2007)

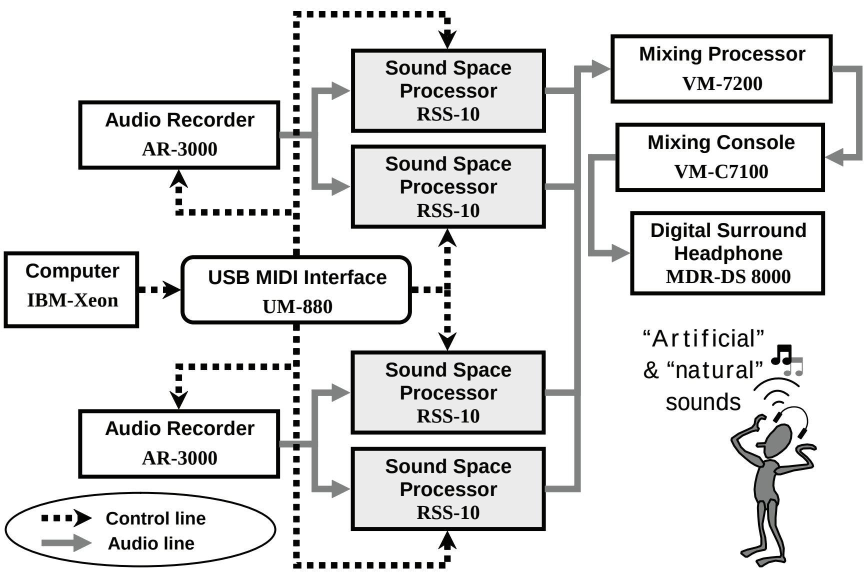
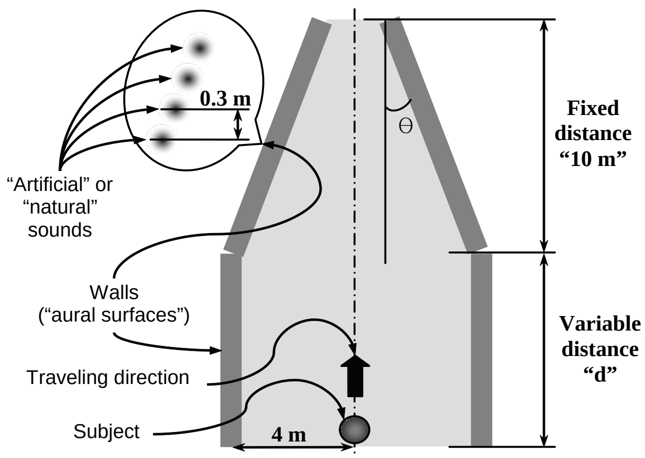


Figure 9. Upper view of created virtual course for experiments (Nomura et al., 2007).



"Artificial" sounds are generated by digitizing sound used by vOICe Learning Edition (Jones, 2004) with no analogs in everyday listening as source. On the other hand, the everyday world sounds of fan noise captured by precise microphones are the source of "natural" sounds (Nomura et al., 2007).

Figure 8 shows the schematic overview of devices, their names and their respective models.

Virtual Courses

The 3-D acoustic environment system virtually creates a course formed by nonspeech sound walls, as shown in Figure 9. Subjects' task consists of traveling across this virtual course and perceiving changes in walls (called "aural surfaces") via their auditory system. Reverberation and reflection levels were set to -30 dB for virtual environment considering results of previous work (Shiose et al., 2004).

Virtual courses for experiments are created for different conditions indicated in Figure 9 as follows:

- The piece corresponding to an "aural surface" of the virtual course is formed by four generated nonspeech sounds ("artificial" or "natural").
- The distance "d" corresponds to a part of the virtual course where there is no change in "aural surface." This distance is calculated by:

$$d = v.t, \quad (1)$$

where "v" can be 4 km/h or 12 km/h, and "t" can be 3.0, 3.5, 4.0, 4.5, 5.0, 5.5, or 6.0 s.

- The distance of 10 m corresponds to the part where the aural surface changes to create different geometric shapes of virtual courses for categorization task by subjects.
- Three different geometric shapes of virtual courses are created according to the value of angle "θ," as follows:
  1. Convergent course when "θ" > 0 such as 3, 5, or 30 degrees. In this case, the virtual course gradually becomes narrow for subjects traveling through this environment as shown in Figure 10 (a).
  2. Straight course when "θ" = 0. In this case, the geometric shape of virtual course is unchanged as shown in Figure 10 (b).

3. Divergent course when "θ" < 0 such as -3, -5, or -30 degrees. In this case, the virtual course gradually becomes wide for subjects traveling through this environment as shown in Figure 10 (c).

Experiments

We carried out the experiments with 14 subjects to evaluate their conceptualization performance. The subjects heard two different sounds ("artificial" and "natural") to perceive changes in the geometric shapes of the virtual courses. No visual information was used by subjects in the experiments and all subjects had no audition problems.

During the experiments, the subjects performed the following tasks:

- The subject travels across a straight course in the virtual acoustic environment corresponding to distance "d" in Figure 9, where the period "t" is randomly selected from seven options (3.0, 3.5, 4.0, 4.5, 5.0, 5.5, or 6.0 s).
- After traveling the distance "d" of virtual course, the subject must press a key on the keyboard corresponding to answer for the perceived type of course as follows:

    1. The "J" key corresponds to the convergent course as shown in Figure 10 (a).
    2. The "K" key corresponds to the straight course as shown in Figure 10 (b).
    3. The "L" key corresponds to the divergent course as shown in Figure 10 (c).

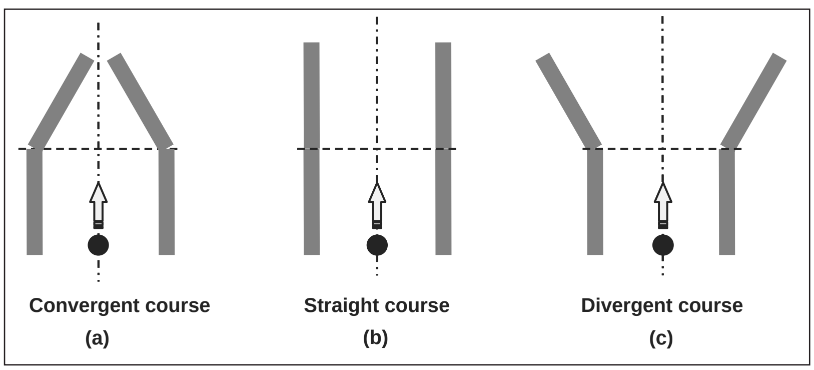
- Each key must be pressed after traveling the distance "d" of virtual course and before arriving at the final of this course. If the subject presses the decision key without certainty then the corresponding result is discarded.

Experiments consisted of subjects performing two experimental stages to analyze the influence of different sounds on their conceptualization performance for spatial information as follows:

- Training stage with two steps, as follows:

    1. In the first step, three courses for each sound ("artificial" or "natural") are virtually created to provide subject's familiarity with the existence of three types of virtual courses as shown in Figure 10. Angles of courses are set at 30, -30, and 0 degrees; traveling speed is set at 12 km/h.
    2. In the second step, the subject is asked to press a key corresponding to the type of perceived course, and after each response, the correct answer is given.

Figure 10. Types of the created virtual courses



- Testing stage is to get experimental results for analyzing the conceptualization performance by subjects hearing different types of sounds. Each subject must try ten sets of experiments. There are two types of sets as follows:

    1. **Set A:** Fourteen different virtual courses corresponding to the combination of two speeds (4 km/h or 12 km/h), and seven angles (-30, -5, -3, 0, 3, 5, or 30 degrees) must be categorized by the subject hearing "natural" sounds.
    2. **Set B:** The unique difference with the set A is that the subject hears "artificial" sounds instead of "natural."

The subjects were divided into two groups (X and Y) and sequences of sets for subjects in each group are as follows:

- Sequence A B B A B A A B A B for group X
- Sequence B A A B A B B A B A for group Y

Figure 11. Conceptualization performance of all subjects traveling at v = 4 km/h

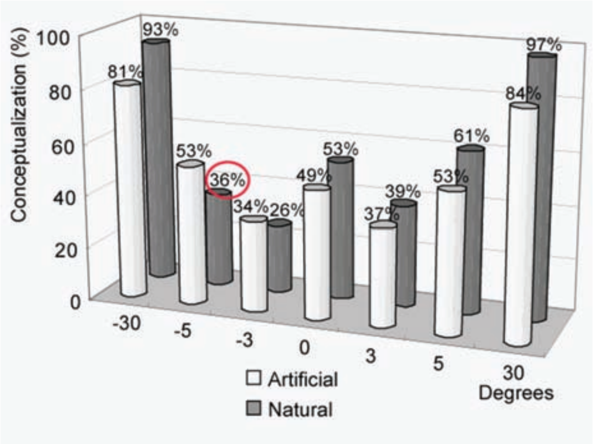


Figure 12. Conceptualization performance of all subjects traveling at v = 12 km/h

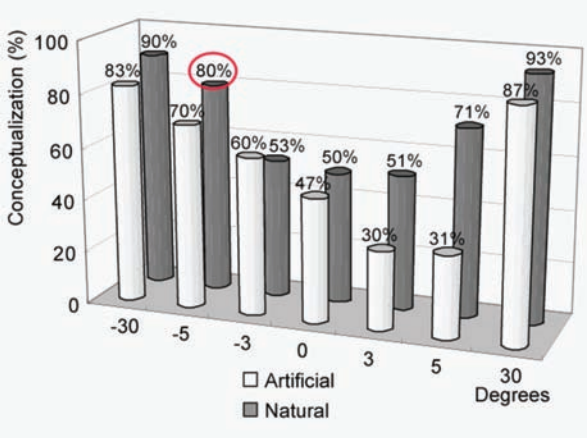


Figure 13. Best conceptualization performance of subjects in group "X"

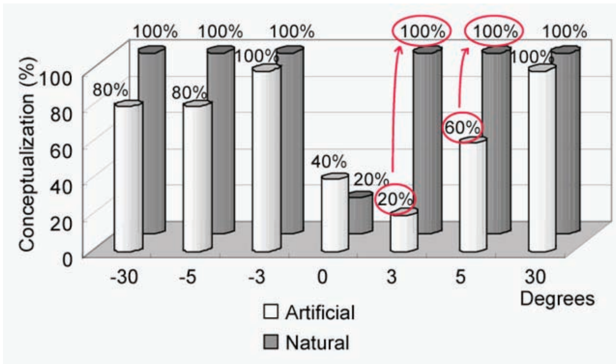
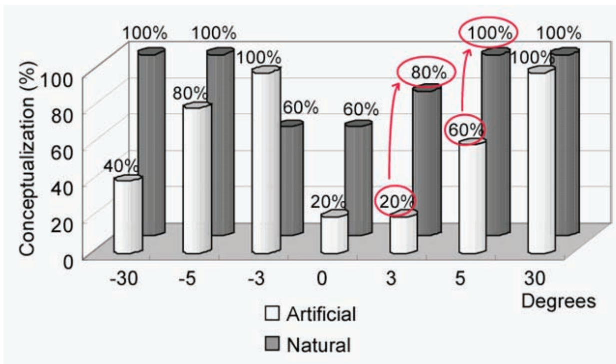


Figure 14. Best conceptualization performance of subjects in group "Y"



Then, each subject tested 140 virtual acoustic courses in the testing stage.

Experimental Results

Figures 11, 12, 13, and 14 show the subjects' abilities to conceptualize spatial information during the test.

Figures 11 and 12 present the average results (%) of all subject's abilities to conceptualize spatial information while navigating at 4 and 12 km/h. The average results are categorized by type of sound ("artificial" or "natural") and course change (degrees).

Figures 13 and 14 present the results for the subjects from groups "X" and "Y" who were best able to conceptualize spatial information while navigating.

We asked subjects which sound ("artificial" or "natural") was the most preferable sound when conceptualizing spatial information. Only three subjects found "artificial" to be preferable.

Discussion

Figures 11 and 12 showed that conceptualization performance with "natural" sounds was better than performance with "artificial" ones in most of the changes on the virtual courses. Also, when traveling speed was set at 12 km/h with the "natural" sounds, we verified a considerable improvement of performance results. For example, a performance of 36% for -5 degrees and 4 km/h in the graph of Figure 11 was improved to 80% for -5 degrees and 12 km/h in the graph of Figure 12.

In Figures 13 and 14, subjects considerably improved their performances traveling across the virtual courses converged by 3 and 5 degrees, and hearing "natural" sounds instead of "artificial." For example, Figure 13 presents the improvement on performance from 20 to 100% with the "natural" sounds in the virtual course converged by 3 degrees and from 60 to 100% with 5 degrees convergence. Also, Figure 14 presents the improvement on performance by similar rates (from 20 to 80% with 3 degrees convergence and from 60 to 100% with 5 degrees convergence).

CONCLUSION

During the experiments, we found that some people have naturally accurate senses for quickly capturing the cues necessary to perceive and localize moving pattern-associated sounds.

The sound localization process we developed using the virtual 3-D acoustic space worked well to support people having difficulties acquiring the necessary perceptual skills for localizing pattern-associated sounds.

The results from the localization experiments showed that the proposed process can improve pattern-associated sound localization ability with the senses, which are considered as perceptual systems in perception analysis, by the ecological psychology approach. Also, we found that sound effects, such as reflection and reverberation added to a moving nonspeech sound source, improved the subjects' abilities to localize sound while navigating through the 3-D acoustic space.

Furthermore, results were better for subjects who only perceived the pattern-associated sounds than for those who tried to recognize or interpret them. In other words, the ability of subjects to localize sound was greatest when the time for perceiving the patterns was short, that is, when the subjects did not try to recognize these patterns.

During the process for conceptualizing spatial information, we evaluated "artificial" and "natural" sounds as nonspeech audio cues.

The experimental results showed that "natural" sounds were better than "artificial" ones for enabling subjects to conceptualize spatial information in different types of virtual acoustic courses. Moreover, the subject's abilities improved when the subjects heard "natural" sounds instead of "artificial" ones during tasks to perceive very small changes in the virtual acoustic courses.

We found that "natural" sounds with conditions similar to those in everyday listening, such as reverberation and reflection, are essential for designing friendly human-environment interfaces based on nonspeech audio. Using such conditions, we can avoid the necessity of using sophisticated devices, which require heavy computational load, to imitate human conceptualization skills for everyday tasks.

The results of our experiments showed that the transfer processes of sound localization and conceptualization skills with a virtual 3-D acoustic space system are possible and useful.

The apparatus used for generating the pattern-associated and "natural" sounds in the virtual 3-D acoustic space is a promising support system. It can be a great alternative to conventional systems for enabling the skill transfer process for users, such as visually impaired persons, who eventually will use wearable devices as media for interacting with the world around them. The visually impaired will enjoy the freedom of accessing everyday facilities without having to depend on the support of conventional "expert systems," such as pattern recognition systems, written language to spoken language converters, and voice synthesizer systems.

Since this research topic is still in its infancy, more exhaustive experiments are necessary to quantify how sensitive a subject (considered as perceptual system) is in accurately localizing pattern-associated sounds and conceptualizing spatial information from nonspeech audio as "everyday listening" skills to perform everyday tasks.

FUTURE TRENDS

Daniel Kish (Roberts, 2006) lost his sight when he was an infant, and he is called the spiritual successor to James Homan. He taught himself to "see" with sonar by clicking his tongue, which enables the conceptualization of spatial information by using acoustics (hearing) similar to how a bat uses echolocation. He is a cofounder and director of World Access for the Blind (Roberts, 2006), an organization that reflects the conviction that blind people can learn to "see" without sight. The "Seeing Without Sight" program teaches blind students to localize and conceptualize objects, such as trees, poles, cars, and bushes without touching them. The students simply click their tongues quietly and understand the patterns of the reflected sound. Kish created the first systematic, comprehensive echolocation training curriculum and designed the first device to enhance sonic echolocation for everyday use.

Because of his success in teaching human echolocation skills to blind people, we believe our concept can be realistically used in future handheld devices.

We hope to contribute to the design of mobile devices that provide benefit of low computational cost by using the concept based on the human skill transfer process we proposed.

Using human echolocation skills, we can develop devices that combine ultrasonic sonar technologies for capturing nonspeech audio cues used during spatial perception and conceptualization tasks. In other words, we believe human interfaces that support the conceptualization of spatial information based on nonspeech audio will contribute significantly to the "Seeing Without Sight" program. Moreover, such interfaces might free blind people from being wired to complex expert systems with conventional interfaces requiring a significant mental workload and hard training.

REFERENCES

Blattner, M. M., Sumikawa, D. A., & Greenberg, R. M. (1989). Earcons and icons: Their structure and common design principles. *Human-Computer Interaction*, *4*(1), 11–44.

Blauert, J. (1997). *Spatial hearing: The psychophysics of human sound localization*. (revised ed.). Cambridge, MA: The MIT Press.

Brewster, S. A., & Walker, V. A. (2000). Non-visual interfaces for wearable computers. In *Proceedings of IEE Workshop on Wearable Computing* (IEE, London) IEE, 00/145.

Brewster, S. A., Wright, P. C., & Edwards, A. D. N. (1993). An evaluation of earcons for use in auditory human-computer interfaces. In *Proceedings of ACM/IFIP INTERCHI'93* (pp. 222–227). Amsterdam: ACM Press, Addison-Wesley.

Bronstad, P. M, Lewis, K., & Slatin, J. (2003). Conveying contextual information using non-speech audio cues reduces workload. In *Proceedings of Technology and Persons with Disabilities Conference*, Los Angeles, CA.

Buxton, W. (1990). Using our ears: An introduction to the use of nonspeech audio cues. In E. Farrell (Ed.), Extracting meaning from complex data: Processing, display, interaction. In *Proceedings of the SPIE*, *1259*, 124–127.

Corey, D. P., & Hudspeth, A. J. (1979). Ionic basis of the receptor potential in a vertebrate hair cell. *Nature*, *281*, 675–677.

Eimas, P. (1975). Auditory and phonetic coding of the cues for speech discrimination of the r-l distinction by young infants. *Perception and Psychophysics*, *18*, 341–347.

Erulkar, S. D. (1972). Comparative aspects of spatial localization of sound. *Physiological Review*, *1952*, 237–260.

Fujiki, N., Riederer, K. A. J, Jousmaki, V., Makela, J. P., & Hari, R. (2002). Human cortical representation of virtual auditory space: Differences between sound azimuth and elevation. *European Journal of Neuroscience*, *16*, 2207–2213.

Gaver, W. W. (1989). The SonicFinder: An interface that uses auditory icons. *Human-Computer Interaction*, *4*(1), 67–94

Gaver, W. W. (1993). What in the world do we hear? An ecological approach to auditory event perception. *Ecological Psychology*, *5*, 1–29.

Gibson, J. J. (1966). *The senses considered as perceptual systems.* Boston: Houghton Mifflin.

Gibson, J. J. (1979). *The ecological approach to visual perception*. Boston: Houghton Mifflin.

Handel, S. (1989). *Listening: An introduction to the perception of auditory events.* Cambridge, MA: The MIT Press.

Haykin, S. (1999). *Neural networks: A comprehensive foundation* (2nd ed.). NJ: Prentice-Hall.

Heine, W. D., & Guski, R. (1991). Listening: The perception of auditory events? An essay review of listening: An introduction to the perception of auditory events by Stephen Handel. *Ecological Psychology, 3*, 263–275.

Jones, W. D. (2004). Sight for sore ears. *IEEE Spectrum, 41*(2), 13-14.

Kallinen, K. (2003). Using sounds to present and manage information in computers. In *Proceedings of the 2003 Informing Science and Information Technology Education Joint Conference*. Informing Science Institute, California (pp. 1031–1043).

Kendall, G. (1991). Visualization by ear: Auditory imagery for scientific visualization and virtual reality. *Computer Music Journal, 15*(4), 70–73.

Mφller, H., Sφrensen, M. F., Hammershφi, D., & Jensen, C. B. (1995). Head-related transfer functions of human subjects. *Journal of the Audio Engineering, 43*(5), 300–321.

Nomura, S., Shiose, T., Kawakami, H., Katai, O., & Yamanaka, K. (2004). A novel "sound visualization" process in virtual 3D space: The human auditory perception analysis by ecological psychology approach. In *Proceedings of the 8th Asia Pacific Symposium on Intelligent and Evolutionary Systems, Cairns, Australia* (pp. 137–149).

Nomura, S., Tsuchinaga, M., Nojima, Y., Shiose, T., Kawakami, H., Katai, O., & Yamanaka, K. (2007). Novel nonspeech tones for conceptualizing spatial information. *Artificial Life and Robotics, 11*, 13-17.

Nomura, S., Yamanaka, K., & Katai, O. (2002). New adaptive methods applied to printed word image binarization. In *Proceedings of the 4th IASTED International Conference Signal and Image Processing* (pp. 288–293). Hawaii .

Nomura, S., Yamanaka, K., Katai, O., & Kawakami, H. (2004). A new method for degraded color image binarization based on adaptive lightning on grayscale versions. *IEICE Trans. on Information and Systems, E87-D*(4), 1012–1020.

Nomura, S., Yamanaka, K., Katai, O., Kawakami, H., & Shiose, T. (2004). Towards a novel "sound visualization" via virtual 3D acoustic environmental media. In *Proceedings of the International Workshop on Intelligent Media Technology for Communicative Intelligence* (pp. 121–124). Warsaw, Poland.

Nomura, S., Yamanaka, K., Katai, O., Kawakami, H., & Shiose, T. (2005). A novel adaptive morphological approach for segmenting characters in degraded images. *Pattern Recognition, 38*(11), 1961–1975.

Plenge, G. (1974). On the differences between localization and lateralization. *Journal Acoustical Society of America, 56*, 944–951.

Roberts, J. (2006). *A sense of the world: How a blind man became history's greatest traveler.* HarperCollins Publishers.

Shaw, B. K., McGowan, R. S., & Turvey, M. T. (1991). An acoustic variable specifying time-to-contact. *Ecological Psychology, 3*, 253–261.

Shiose, T., Ito, K., & Mamada, K. (2004). The development of virtual 3D acoustic environment for training 'perception of crossability'. In *Proceedings of the 9th International Conference on Computers Helping People with Special Needs.* Paris, France.

Shiose, T., Sawaragi, T., Nakajima, A., & Ishihara, H. (2004). Design of interactive skill-transfer agent from a viewpoint of ecological psychology. *International Journal of Human-Computer Interaction, 17*(1), 69–86.

Vanderveer, N. J. (1979). *Ecological acoustics: Human perception of environmental sounds.* Doctoral Thesis, Dissertation Abstracts International, 40/09B, 4543 (University Microfilms No. 8004002).

Wall, S. A., & Brewster, S. A. (2006). Tac-tiles: Multimodal pie charts for visually impaired users. *4th Nordic Conference on Human-Computer Interaction, ACM International Conference Proceeding Series 189*, 9–18.

Warren, W. H., & Verbrugge, R. R. (1984). Auditory perception of breaking and bouncing events: A case study in ecological acoustics. *Journal of Experimental Psychology: Human Perception and Performance, 10*(5), 704–712.

Wise, G. B. (1993). *Beyond multimedia: An architecture for true multimodal interfaces.* A Research Proposal. Rensselaer Polytechnic Institute, New York.

Yeung, E. S. (1980). Pattern recognition by audio representation of multivariate analytical data. *Analytical Chemistry, 52*, 1120–1123.

KEY TERMS

"Artificial" Sounds: Have no analogs in everyday listening and require extensive trials before actually using the interface

Conceptualization: Task for visually impaired persons to construct categories by capturing spatial information cues in nonspeech audio

Everyday Listening: The experience of listening to events rather than sounds, that is, skills relied upon in everyday tasks such as driving and crossing the street

"Natural" Sounds: Consider everyday listening to provide friendly information for users without hard cross-modal training

Nonspeech Audio: Has benefits such as the increase of information communicated to the user, the reduction of information received through the visual channel, the performance improvement by sharing information across different sensory modalities

Pattern-Associated Sound: Represents the most appropriate sound event to be perceived even by visually impaired persons

Sound Localization: Learning process for visually impaired persons considered as perceptual systems to localize pattern-associated sounds in a virtual 3-D acoustic space

Spatial Information: Such information as size, shape, and texture of objects from the pattern of the reflected and reverberated sounds

Chapter XLI
A Navigational Aid for Blind Pedestrians Designed with User- and Activity-Centered Approaches

Florence Gaunet
Laboratoire Eco-Anthropologie et Ethnobiologie UMR 5145, CNRS, France

Xavier Briffault
CESAMES UMR 8136, Université René-Descartes Paris V, France

ABSTRACT

The two-fold aim of this chapter is to present the design process of an interface for a mobile navigational aid for blind pedestrians and a set of rules for producing route descriptions for these users, as well as the methodology used to develop them, rooted in a user- and activity-centered approach. We first present the state of the art of wearable verbal navigational aids and what might still be lacking in their conception, and propose a reusable user- and activity-centered approach designed to complement already existing and future systems. Case studies fitting into this approach are next presented: route descriptions produced by blind pedestrians were analyzed; the production rules were extracted and tested in urban areas. Results reveal these rules, the specific database features, the required user profiles, and the precision of localization necessary for assisting blind pedestrians' wayfinding in urban areas. Finally, future trends in mobile guiding tools for the visually impaired are examined.

INTRODUCTION

Imagine being blind and trying to find your way around a city you have never visited before; that can even be challenging for a sighted person. Wayfinding (the tactics and strategies that guide traveling) is particularly difficult for blind people to achieve, even in our highly structured modern cities. Indeed, though the vast majority of blind people live in urban environments, most of them are prevented from traveling autonomously in cities. This is especially true in unfamiliar areas due to the lack of adequate and accessible information to guide them (Bentzen, 1997a; Golledge, Klatzky, Loomis, Spiegel, & Tietz, 1998). Thought a few wearable navigational aids exists, their usability by blind individuals in real-life situations has not been studied in detail, and uphill field studies have not been performed to identify information necessary to design proper mobile way-finding aids.

Our goal in this chapter is to present the state of the art of wearable and auditory navigational aids (devices and research projects), identify what might still be lacking in their conception, to propose a user- and activity-centered approach designed to get data necessary to complement already existing and future systems, and finally to provide case studies fitting into this approach. The two-fold objective of this chapter is to present the set of rules for producing verbal instructions and information, and the methodology we have developed for designing the system and the set of rules, based on case studies.

NAVIGATIONAL AIDS FOR VISUALLY IMPAIRED PEDESTRIANS

The idea of developing a wearable computerized guidance system using satellite-based localization (GPS – global positioning system or Galileo) to assist the navigation of visually impaired goes back two decades (Collins, 1985). Such a system is not intended to provide visually impaired persons with detailed information about the immediate environment (e.g., obstacles), but to provide way-finding information and instructions. Thus, the traveler still has to rely on the long cane, the guide dog, or ultrasonic sensing devices for this type of information.

The first verbal and wearable guidance system for visually impaired people was designed back in 1989 (Brusnighan, Strauss, Floyd, & Wheeler, 1989). More recently, a prototype was designed by the MoBIC consortium in England (Petrie, Johnson, Strothotte, Raab, Fritzand, & Michel, 1996). Since then, there have been a multitude of research projects investigating GPS-based navigation systems for visually impaired travelers (e.g., Fruchterman, 1996; Helal, Moore, & Ramachandran, 2001; Holland, Morse, & Gedenryd, 2002; LaPierre, 1998; Makino, Ishii, & Nakashizuka, 1996). A wireless prototype GPS system for the visually impaired is also under development in Europe in the TORMES project by a Spanish company called GMV Sistemas and the ONCE (the National Organization of Spanish Blind people). It is a handheld device using EGNOS satellite technology, created by the European Space Agency (ESA).

These systems are most often built around a technical architecture composed of a geographic information system (GIS), coupled with path computation software, a localization system (GPS alone, or GPS plus a dead reckoning system), a wearable computer, and a human-computer interface, often based on speech synthesis (Gaunet & Briffault, 2001, 2002; Golledge et al., 1998; Loomis, Golledge, & Klatzky, 2001). An alternative type of navigational aid signals the position of landmarks to visually impaired travelers through stereo headphones. It provides a three-dimensional spatial virtual acoustic display for landmark localization, and labels and attributes of landmarks with speech synthesis (see, for example, the personal guidance system designed by the University of California, Santa Barbara research group – Loomis et al., 2001, and the System for Wearable Audio Navigation SWAN developed by Georgia Tech – Walker & Lindsay, 2006).

Though most of these systems have been validated from a strictly technical point of view (i.e., localization, speech synthesis, integration of components, etc.) few effective commercial systems are available. These do not use tactile or 3-D-sound spatial display but voice messages, suggesting at least some technological and usability issues with the two former modalities (e.g., wearability of an electronic tactile map and headphones for conveying binaural sounds for a satisfactory spatialization and surroundings perception). Moreover, it is well known that blind pedestrians who are expert to a path can efficiently verbally guide pedestrians naive to that path (Gaunet & Briffault, 2005), the latter using their cellular phone. Currently, visually impaired consumers in the U.S. can choose between three GPS products based on verbal route descriptions: the Trekker produced by HumanWare, a GPS system that connects to a personal data assistant (HumanWare Web site); a GPS system from the Sendero Group that connects to a BrailleNote portable note taker (Fruchterman, 1996; May, 2002; Sendero Web site) and a GPS system for PAC Mate produced by Freedom Scientific (Freedom Scientific Web site). In France, the main company delivering tools for the visually impaired has recently proposed the Wayfinder Navigator device, a navigation tool designed for sighted people (Wayfinder Web site). Yet, the usability of these devices by visually impaired individuals in real-life situations remains questionable. Indeed, these devices are not yet part of the everyday life of visually impaired persons, unlike Jaws, the widespread screen reader. Indeed, detailed scientific user studies are not available, and the research reports quoted here (as well as the few others that exist), along with the very low sales, may indicate that some issues remain to be solved regarding these systems; these may be related to the design of the guidance instructions or to the too limited number of functions for instance. In fact, none of these systems is based on a precise analysis of the visually impaired person's way-finding activity, nor do they take into account the guidance methods actually used by visually impaired people. For instance, user profiles, including the different types of visual impairments or the methods used by pedestrians to detect obstacles have not yet been taken into account. The features of future mobility devices must take into consideration the abilities, needs, and wishes of visually impaired people. Too many times, devices have been proposed by technology-oriented researchers and developers who often show little awareness of the needs and the perceptual and cognitive abilities of the intended users (Loomis, 2005). User- and activity-centered design is a different approach, starting with the needs and characteristics of the population to be served and with the requirements related to the latter perceptual and cognitive abilities in the intended context.

A USER- AND ACTIVITY-CENTERED APPROACH FOR DESIGNING A NAVIGATIONAL AID FOR VISUALLY IMPAIRED PEDESTRIANS

The user-centered approach consists in bootstrapping a user model from existing data, often obtained in experimental settings originating from the domains of cognitive psychology and human factors research, and in incrementally updating it with an activity-centered approach developed in the ecological context of use of the future device. Figure 1 shows an overview of the specification process. The left column shows the concrete steps that were used (see the section describing the field studies "Specification process for a verbal guidance system for blind pedestrians"); the middle column shows the nature of the specifications gathered during the case studies; the last column shows the progress toward identifying the guidance functions that were obtained. Finally, the steps corresponding to both the user-centered and the activity-centered approach are highlighted on the right side of the figure.

This specification process is derived from a more generic one, presented in Kovacs, Gaunet, and Briffault (2004) in which dedicated techniques (*observation*, *wizard of Oz*, *questionnaire*…) are used at different steps of an incremental conception and validation process, see Figure 2; these techniques are described in detail in Kovacs et al. (2004).

Figure 1. Overview of the specification process (steps are referenced in the following section).

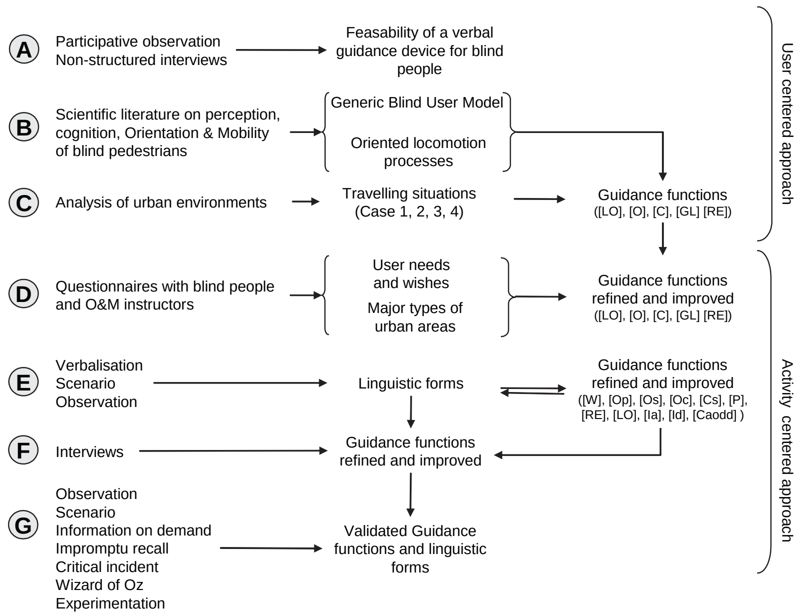
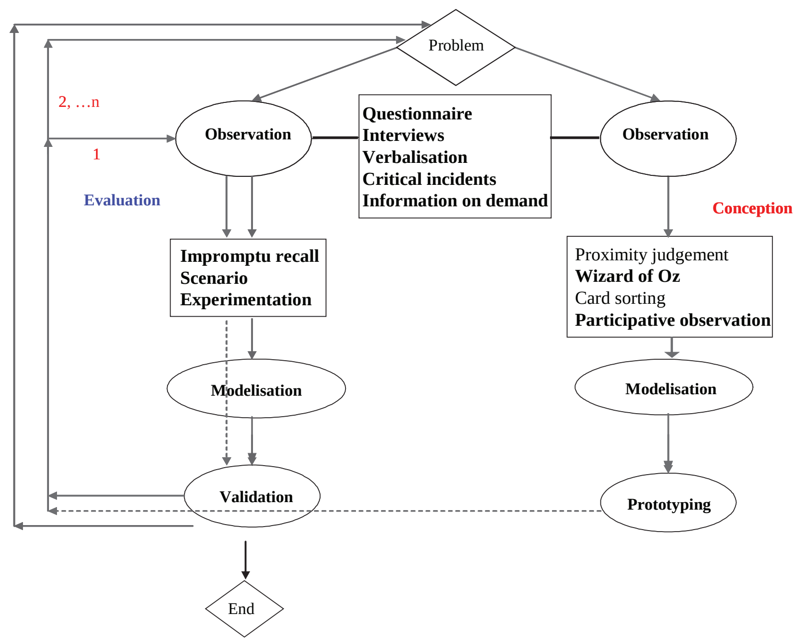


Figure 2. The different techniques used in the specification process (Adapted from Kovacs, Gaunet, & Briffault, 2004)



SPECIFICATION PROCESS FOR A VERBAL GUIDANCE SYSTEM FOR BLIND PEDESTRIANS: CASE STUDIES

The results presented here concern a verbal navigational aid for totally blind pedestrians, a subpopulation of the visually impaired (Gaunet, 2006; Gaunet & Briffault, 2005). Because visual impairments are of various types (no vision, peripheral or central vision deficiency, blurred vision for instance), studying how to guide the person who is blind have to be considered as specifying a user profile. The design and validation of a set of rules for producing verbal instructions and information to guide blind pedestrians in unknown urban environments are presented here in concrete terms. Such a set of rules, which relies on the cognitive abilities and the sensori-motor information available to blind pedestrians, is an absolute prerequisite for the design of wearable navigational aids.

The key points to be identified for the design of a functional system (Gaunet & Briffault, 2001, 2002) are:

- The set of rules for producing verbal instructions and information as a function of the characteristics of the urban environment to be traveled, the required linguistic forms, and guidance functions (verbal information and instructions)
- The components of the urban environment to be included in the geographical database
- The optimal distance at which guidance functions should be triggered before the required action is to occur
- The localization precision needed for effective assistance

The benefits of the approach proposed here, compared with those implemented in the past, is that firstly it is based on ergonomic research, that is, on an analysis of how blind pedestrians actually perceive the urban environment during oriented locomotion and the traveling activity of end user. Secondly, it is to be conducted prior to the design and manufacture of the technical aid.

User-Centered Approach

The aim of this part of the specification process is to specify a user-model, which includes a description of the abilities of the future user.

Nonstructured interviews[1 2] of blind people, as well as many years of *participative observation* (see Figure 1, A) of blind peoples' everyday activities by one of the authors, revealed that blind pedestrians can, and do, successfully guide each other over cellular phones (Gaunet & Briffault, 2005). Thus, strictly verbal directional instructions and descriptions, without visual knowledge of the described environment by the guide, are sufficient to efficiently guide a blind person through an unfamiliar environment.

Finding a solution to the navigational problem (the aggregate task of way finding and traveling) requires first of all an in-depth understanding of the accuracy at which the nonvisual senses of blind people perceive environmental features and localize objects and themselves, as well as the nature of the urban spatial structure to be traversed and perceived, and, finally, the mobility and way finding abilities of blind pedestrians.

The literature on perception, cognition, and orientation and mobility (O&M) of blind pedestrians (see Figure 1, B) shows that the information conveyed by the remaining senses does not provide accurate perception of space and events in urban areas (Foulke, 1982; Guth & Rieser, 1997; Jansson, 2000; Rieser, 1999; Thinus-Blanc & Gaunet, 1997). Indeed, the blind person perceives a spatial array that could be metaphorically described as "blurred." Objects can be detected by variations of the ambient sound field as one moves around or makes noise, and certain objects, or categories of objects, can be identified by the sounds they make (Hughes, 1989). However, they lack the ability to recognize shape, size, texture, and material with precision. The accuracy of recognition as well is obviously very poor. As stated by Jansson (2000), sounds may roughly indicate the location of important landmarks. For example, the sound of traffic (and more generally the environmental flow) indicates the location and orientation of a street (Ashmead, Wall, Eaton, Ebinger, Snook Hill, Guth, & Yang, 1998). The reverberation of sound produced by some objects creates a "wall of sound" on one side of the traveler, which can be used to travel a straight line (Ashmead & Wall, 1999). Spectral variations in the ambient sound field coming from roundabouts provide rough information about their spatial structure (Wiener, Lawson, Naghshineh, Brown, Bischoff, & Toth, 1997). The sound of a vehicle gives information about its movement (whether it is approaching or moving away and at what speed), and sounds reflected by objects enable the detection of obstacles (Wiener & Lawson, 1997). The current development of external earphones is thus a key factor for the development of verbal navigational aids (see for instance the hearing device developed in SWAN – see the SWAN Web site). Moreover, elements of the locomotor system (the vestibular system, joints, muscles, and tendons) provide information on the position of the limbs and of the body itself with respect to the environment (spatial updating), as well as on linear and angular head and body movements (path integration) (Loomis et al., 2001; Millar, 1994; Rieser, 1999). The haptic system is used to touch objects and recognize environmental features (Hatwell, Strerri, & Gentaz, 2000). The spatial range of these systems is obviously limited by the length of arms and legs, with the possible extension provided by the long cane or the dog's harness (Long, 2001; Millar, 1994). Temporal updating is essentially sequential, as objects are examined one by one, and is thus slow. Cutaneous information, provided by the skin, can also be used in various locomotor and haptic situations. Integrating the sensation of the sun's heat with the time of the day provides clues about the geographical direction (Foulke, 1982; Guth & Rieser, 1997). The appearance and disappearance of sunshine or gusts of wind as one rounds a street corner can be felt by the skin and provide information about crossroads. The smell of a place can serve as a landmark.

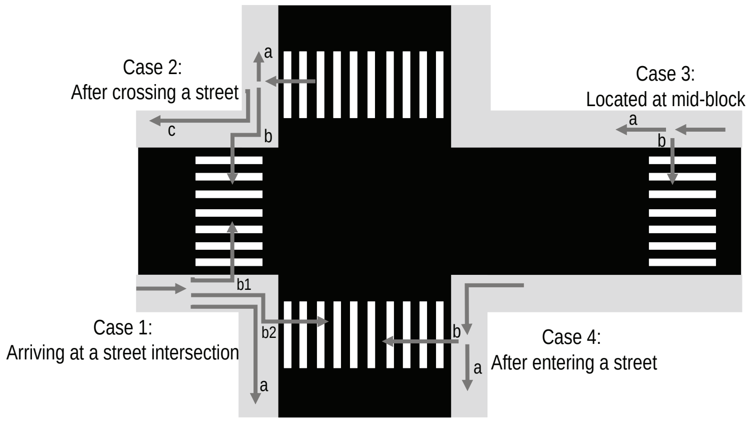
The spatial array perceived by a blind person is thus of poor perceptual acuity, and only provides a basis for low-reliability oriented locomotion. According to the author's participative experience and that of O&M instructors and blind people themselves, collisions with obstacles, slowness, and the necessity of complicated strategies to find targets are often reported. Consequently, the verbal information provided for guidance should include clues about the spatial architecture array so that pedestrians can build local and temporary mental representations that fit their specific spatial representation of the environment for decision making and to control their travel (Golledge, 1999; Rieser, 1999; Thinus-Blanc & Gaunet, 1997). Though guide dogs are useful for many of these purposes and can ease locomotion, there is no evidence in the literature to suggest that cane and dog users find their way differently (Bentzen, 1997b).

Integrating the results on the abilities of the blind pedestrians presented previously, with an analysis of the main architectural and urban items blind pedestrians encounter (see Figure 1, C), allowed Gaunet and Briffault (2005) to establish a taxonomy of the activities performed by blind pedestrians, and the information required to guide them along paths according to the architectural structure of urban areas and pedestrian activity. Four different cases, corresponding to specific traveling situations, each with its specific guidance functions (presented below) may be encountered (see Figure 3):

- Arriving at a street intersection (Case 1). The pedestrian either turns onto an adjacent street (case 1a), crosses the street on his/her side (case 1b1), or crosses the street in front of him/her (case 1b2),
- After crossing a street (Case 2). The pedestrian either turns onto a street (case 2a), crosses a crosswalk on the side (case 2b), or walks forward along a street (case 2c),
- Located at mid-block (Case 3). The pedestrian may either walk along the street (case 3a) or cross the street (case 3b),
- After entering a street (Case 4). The pedestrian may either walk forward (case 4a) or cross a street on his/her side (case 4b).

These are the kinds of situations (choice points for a new action to be performed) that need to be taken into account by the navigational aid.

Figure 3. The four different traveling cases for blind pedestrians in urban areas (case 1, 2, 3, 4) (Adapted from Gaunet and Briffault, 2005)



Along paths, and according to where the pedestrian is located along the path, each of these four traveling situations can be identified through the path computation process performed either by a (human) guide or the wearable guidance device. Because of the impaired cognitive and perceptual anticipation abilities of blind pedestrians during oriented locomotion, guidance functions should be provided from five to ten meters prior the action they prescribe.

A synthesis of the perceptual and environmental information available to blind pedestrians, and of the processes of oriented locomotion (Foulke, 1982; Jansson, 2000; Long, 2001), indicates that achieving travel requires six kinds of route descriptions, the so-called guidance functions. *Information* guidance functions are "Localization and Orientation [LO]," which provides information on a pedestrian's location (e.g., address), and orientation (e.g., oriented with the flow of traffic), and "Goal Location [GL]," which provides information about the goal location (e.g., distance and heading, address…). *Instruction* guidance functions are "Orientation [O]," which provides instructions about orienting toward a target (e.g., *taking a street to the right/left or at mid-block*), "Crossing [C]," which provides instructions about how to cross a street, "Route Ending [RE]," which informs that the goal has been reached, and "Progress [P]," which tells the user to continue walking forward. [LO], [GL] [O], [C], and [RE] guidance functions apply in all four traveling situations. [P] applies to traveling situations 2, 3, and 4.

Activity-Centered Approach

To refine the guidance model presented, we gave a *questionnaire*[3] to both blind people and to O&M instructors (see Figure 1, D). Here are examples of three of the open questions asked (Gaunet & Briffault, 2005):

- How do you get through unfamiliar environments?
- What information do you need to be guided in unfamiliar simple and structured environments?

  After the experimenter explained the set of guidance functions for assisting way-finding to the interviewees, they were asked:
- To which extent is describing a path with such guidance functions scripts a correct way to represent blind way-finding activity when traveling in unfamiliar environments?

Overall, the answers allowed going beyond previous data gathered through the user-centered approach concerning the four keys to the design of a functional system.

First, the information gathered confirms that it is possible to verbally guide blind pedestrians in unknown simple and structured urban areas without any localization device using only a cellular phone, but that guidance may fail when sidewalks are too wide or not bordered by walls and curbs.

Second, wherever they are in the environment, the blind persons expect instructions telling them whether to go straight on (a few meters or to continue on the sidewalk where they are), to turn in place (in order to cross the adjacent street or to turn left or right onto a new street or to reorient after a crosswalk, either in terms of degrees or left and right), or to cross a crosswalk (described in terms of distance, direction, and configuration, or in terms of actions to perform). Some subjects required additional information, such as their location (address) along a path, where salient environmental features were with respect to them, and when to stop. Two of the blind participants were not able to clarify their needs.

The answers to the last question indicated that precise localization could significantly improve way finding at intersections for both cane and dog users, and to aid cane users to recover from disorientation. Except in walking speed and in the difficulty that cane users have finding crosswalks, dog and cane users do not seem to differ. Moreover, though overall the respondents expressed that the way-finding taxonomy/schematization matched their current way-finding activity in an unfamiliar environment, we learned new things. For instance, information about where the blind pedestrian is currently located ([LO]) and where the goal is ([GL]) are needed not only at the start of travel, but also during the travel in order to get help, if necessary. The path end-point needs to be given explicitly ([RE]), but the precise linguistic content of this instruction was not specified in the interviews. Regarding instructions about following straight-line paths ([P]), we found that this instruction needs to be given in reference to the type of linear feature to be followed (e.g., a curb or a wall). However, this kind of information is not available in most current geographical databases, and must thus be replaced by available features (e.g., progress along a street). Finally, the orientation function [O] has to specify the orientation in terms of the heading to be followed (in hours or degrees) or turn left/right for orientation along a street, towards a crosswalk, and after a crosswalk, and the precise location of a crosswalk ([C]) should be provided. We observed that both [C] and [O] include directional instruction for finding crosswalks.

Gaunet and Briffault (2005) have finally determined that two major types of urban areas are relevant to the design of rules for producing verbal instructions and information, namely simple and complex urban environments. The present work adopts this two-fold categorization. Simple urban environments are spatially structured areas (i.e., streets bordered by narrow sidewalks, walls, or fences, with cross- and T-intersections, crosswalks perpendicular to sidewalks, one-step crossing, and roads less than 15-m-wide). Most visually impaired people live and travel in such environments. Complex urban environments are unstructured areas, such as open areas, campuses, or major roundabouts, which are usually avoided by blind pedestrians. In these areas, traveling situations require following wide sidewalks (more than 15 m-wide), crossing, entering or traversing pedestrian streets, recovering orientation at the ends of pedestrian streets, using crosswalks not perpendicular to the sidewalk, traversing two-step or nonaligned crosswalks not perpendicular to the sidewalk, crossing wide roads, and traversing large and complex roundabouts (following wide sidewalks around the perimeter).

Thus, the four key points for the design of a functional system noted in the introduction have been refined and enlarged.

Given that blind pedestrians who are expert to a path can verbally guide pedestrians who are naive to that path, we have applied the *verbalization* technique[4] to blind pedestrians (see Figure 1, E) to develop lists of verbal instructions for specific paths. We then extracted the underlying rules for producing verbal instructions and information (Gaunet & Briffault, 2005). This technique was combined with the *scenario* technique[5] in which specific paths were chosen by the blind participants, and the *observation* technique[6] to record the verbal instructions and the actions performed at the locations where they were provided (see Figure 1, E).

In this study (Gaunet & Briffault, 2005), a blind traveler familiar to a path provided verbal instructions to guide a sighted traveler, whom he/she was to consider as blind, through an unfamiliar structured urban area. The blind guide, using both his/her direct perception of the environment and spatial representation of the area, needed first to identify the location of the sighted experimenter, and then to produce a concise description of the route (Golledge, 1999). On average, the distance between the blind guide and the experimenter was less than five meters. The verbal descriptions and where they were provided along the route were recorded.

The [LO], [C], [P], [O], [GL], and [RE] guidance functions were then assigned to the recorded linguistic forms. Because some linguistic forms did not match certain functions, new functions were determined and some previously defined functions were refined (i.e., guidance subfunctions were determined). This assignment allowed not only refining the guidance functions, but also the determination of the linguistic forms used by blind pedestrians to guide each other. The four traveling situations were used to establish a correspondence between guidance functions and the actual activity in the environment.

Further, immediately after the test, the experimenter asked the participant which guidance functions he/she thought would be useful to guide someone using a localized wearable verbal guidance aid, in an unfamiliar simple and structured environment. These semistructured *interviews* (Figure 1, F) revealed the need for:

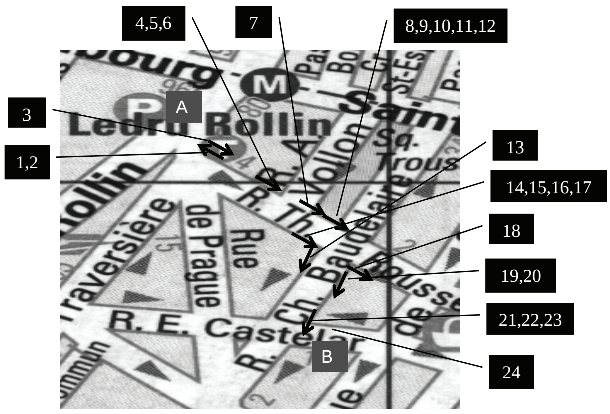
- As much environmental information as possible corresponding to the [LO] function (street name—and street number if possible, the side of the sidewalk, the number of lanes, the orientation of the pedestrian, and the orientation of nearby cars)
- The orientation of crosswalks with respect to the sidewalk (if not perpendicular), the description of the crosswalk if more than one step (number of steps, presence of island, length, width) corresponding to a refinement of the [C] guidance function [Cd], however, in simple and structured areas this refined guidance function is not necessary because crosswalks are single step and perpendicular to sidewalk
- The orientation to take after crossing, now identified as [Oc]
- Descriptions of intersections, now identified as [Id]: no preference was expressed for either deictic or clock mode for describing street positions, but this description needs to be provided within 30 meters of the intersection.

Rules for producing verbal instructions and information were thus refined and linguistic forms recorded.

Route descriptions of unfamiliar paths were then developed based on these results (see an example of an unfamiliar path for a blind pedestrian on Figure 4, and the corresponding instructions that were provided on Table 1). These descriptions were tested for paths unfamiliar to the blind pedestrians using a combination of the *scenario*, *information-on-demand*[7], and *observation* techniques (Figure 1, G). The efficiency of the rules (and of the linguistic forms) of the collected route descriptions was successful: very few errors or hesitations were recorded.

The final results of Gaunet and Briffault's (2005) case studies are that rules for producing verbal instructions and information for blind travelers with five meters' localization (without any orientation detection device) depend on a number of stable features: streets (location, name, length, number, number of lanes, and direction of vehicle traffic), sidewalks (located right or left, possibly width), intersections (name, number of streets, and street orientation), and crosswalks (location, number of steps, possibly the orientation and the configuration of crosswalks). The location and characteristics of traffic lights, and metric guidance are important, but not essential in simple and structured areas. Indeed, blind pedestrians

Figure 4. Example of an unfamiliar path. Arrows depict the location and heading of the pedestrian and where the verbal descriptions were provided. Numbers correspond to the verbal information and instructions provided to the pedestrian (see Table 1). A: Start, B: Arrival (Adapted from Gaunet and Briffault, 2005).



have learned to cross streets without knowing whether or not there are traffic lights. However, this information can be a major plus. Metric guidance can considerably increase the range of achievable travels, and also provides confirmation of position, an essential function for blind pedestrians. On the contrary, localization within 5-10 meters makes it possible to solve essential localization problems (e.g., localizing crosswalks, on which side of the sidewalk one is standing) and to have the system confirming whether a completed action is correct (e.g., for street crossing or confirmation of direction).

Moreover, the four cases of traveling situations define a geographical functional network for blind travelers, in which there are four different types of locations or choice points. At each location, a number of possible guidance functions (and associated linguistic forms) apply according to a number of traveling situations (1a, 1b1, 1b2, 2a, 2b, 2c, 3a, 3b, 4a, 4b). Guidance functions (see Table 2) are composed of route instructions ([W], [Op], [Os], [Oc], [Cs], [P], [RE]) delivered 5-10 meters prior to each action to be performed, and environmental information ([LO], [Ia], [Id], [Caodd]) is provided on the spot (within 5 meters) (except that in case [Ia] the information should be provided 30 meters before the intersection). Following this set of case studies, the rules for producing verbal instructions and information were considerably refined with respect to the specifications given by the user-centered approach alone. Table 2 displays the guidance functions that were finally defined (detailed linguistic forms and formalization of guidance functions are provided in Gaunet and Briffault, 2005).

We next validated these production rules in the two types of urban areas previously defined (Gaunet, 2006). For this purpose, the *scenario, information-on-demand,* and *wizard of Oz*[8] techniques were combined (see Figure 1, G). Itinerary scenarios were defined and tested for the two types of urban environments (simple and structured, and complex and unstructured) using an unmasked *wizard of Oz* setting. The blind participants were guided by a sighted experimenter walking few meters behind them, simulating the future navigation aid device by strictly applying the rules for producing verbal instructions and information. Pedestrians asked for guidance instructions when they judged that the action corresponding to the previous instruction had been completed. On the basis of the previous results, instructions for the two types of areas were developed. In the complex and unstructured case, the rules were modified following the requirements expressed in the

Table 1. Linguistic forms provided to a blind traveler, traveling situations and guidance functions, example of an unfamiliar path (Adapted from Gaunet and Briffault, 2005)

Linguistic Forms	Traveling situations	Guidance functions
1- You are on Roussel Street, on the right(side) sidewalk, at #14, it is a one-way street and you are in the opposite direction of the cars.	Case 3	LO: Localization/Orientation
2- Turn 180°.	Case 3a	Op: Orientation in Place
3- Walk forward on Roussel Street.	Case 3a	P: Progression
4- You arrive at a three-branch intersection.	Case 1	Ia: Intersection Announcement
5- At 6 hrs, there is Roussel Street; at 12 hrs, there is Roussel Street and at 8 hrs, there is Vollon Street	Case 1	Id: Intersection Description
6- Cross Vollon Street.	Case 1b2	Ca: Crossing Announcement
7- After crossing Vollon Street, walk straight on Roussel Street.	Case 2c	Oc: Orientation after Crossing
8- You arrive at a four-branch intersection.	Case 1	Ia: Intersection Announcement
9- At 8 hrs, there is Baudelaire Street, at 12 hrs, there is Roussel Street; at 2 hrs, there is Baudelaire Street; at 6 hrs, there is Roussel Street.	Case 1	Id: Intersection Description
10- You must cross Roussel Street.	Case 1b1	Ca: Crossing Announcement
11- Search for Roussel crosswalk.	Case 1b1	Cs: Crossing Search
12- Cross the street.	Case 1b1	Ca: Crossing Announcement
13- After crossing Roussel Street, orient yourself so as to have Roussel Street on your left.	Case 2	Oc: Orientation after Crossing
14- You are on the right sidewalk of Roussel Street. You are in the same direction as the cars.	Case 2	LO: Localization/Orientation
15- Cross Baudelaire Street.	Case 2b	Ca: Crossing Announcement
16- Search for the Baudelaire crosswalk.	Case 2b	Cs: Crossing Search
17- Cross the street.	Case 2b	Ca: Crossing Announcement
18- After crossing Baudelaire Street, orient yourself so as to have Baudelaire street on your right.	Case 2	Oc: Orientation after Crossing
19- You are on the left sidewalk of Baudelaire Street. You are traveling in the same direction as the cars.	Case 2	LO: Localization/Orientation
20- Walk forward in Baudelaire Street.	Case 2c	P: Progression
21- You arrive at a four-branch intersection.	Case 1	Ia: Intersection Announcement
22- At 6 hrs there is Baudelaire Street; at 8 hrs, there is Castelar Street; at 12 hrs, there is Baudelaire Street and at 2 hrs, there is Castelar Street.	Case 1	Id: Intersection Description
23- Turn left into Castelar Street.	Case 1a	Os: Orientation on a Street
24- You're there.	Case 4	RE : Route Ending

study by Gaunet and Briffault (2005). The same rules were used for both cane and dog users. The earlier finding that walking speed and crosswalk localization differed among these two types of users led us to analyze way-finding performance according to the technical aid used for traveling. The *experimentation* technique[9] was thus used to compare the performances of the four groups defined by two factors (the two types of blind users and two types of urban areas), each composed

Table 2. List and descriptions of guidance functions

Warning [W]	This is information that signals deviation from the expected path. After one minute or 60 meters walking in the wrong direction, the magnitude of deviation is computed, which – depending on its value, may trigger the [LO] function.
Localization and Orientation [LO]	A detailed description of the pedestrian's location (street, street number of specific building(s), number of lanes, sidewalk side and intersecting streets when located on a corner) and orientation (direction of vehicle traffic with respect to pedestrian orientation, direction of the pedestrian with respect to objects in the database) is provided on demand. This information is also provided after triggering [W].
Intersection Announcement [Ia]	This instruction warns of an intersection within 30 meters and gives its name, if any.
Intersection Description [Id]	This instruction gives the number of streets, their names, the clockwise or deictic description of the lane directions, in reference to the location of the pedestrians.
Orientation in Place [Op]	The rotation to be performed is provided at mid-block and when the pedestrian has traveled more than 1 minute off trajectory. In the latter case, the [Op] verbal statement is preceded by [W].
Orientation on a Street [Os]	The instruction to walk into a street on the same sidewalk must be given before entering a street.
Orientation after Crossing [Oc]	This instruction gives the orientation to be taken after crossing.
Crossing Announcement, Orientation, Distance and Description [Caodd]	The announcement, the orientation, the distance and the description (one step in this case) of the next crosswalk to be crossed is given.
Crosswalk Search [Cs]	This function indicates that the next crosswalk must be searched for, once the pedestrian arrives on the spot.
Progress [P]	The traveling progress along a street or the distance to be traveled is explicitly given.
Route Ending [RE]	Achievement of the destination is indicated.

with two modalities (cane vs. dog, and simple and structured vs. complex and unstructured areas). When participants became lost, they were interrupted and asked to describe their thinking process (*impromptu recall* technique[10]) and to report and replay their actions prior to becoming lost (*critical incident* technique[11]), though these situations seldom occurred.

The evaluation of the rules for producing verbal instructions and information was undoubtedly successful for the tested population: very few hesitations or errors occurred during path way finding; moreover, the [LO] function was requested very few times. This result confirms the previous data obtained by Gaunet and Briffault (2005) for simple and structured areas, and also validates the changes performed for the rules applied to complex and unstructured areas.

Furthermore, in contrast with our previous results, this study showed that the verbal instructions work best for dog users traveling in simple and structured areas, second best for dog users traveling in complex and unstructured areas, next for cane user way finding in simple and structured areas, and lastly for cane users in complex and unstructured areas. Thus, the user profile may affect the precision of localization and the descriptions of the location of items in space. Dog users appear to benefit more than cane users from these production rules. Travel times were longer for cane users and their hesitations and errors suggest that their slowness is due to the difficulty they have finding crosswalks, recovering from crosswalks and getting oriented on wide sidewalks or in pedestrian streets. This suggests that the two categories of users differ more in finding cues than in the way-finding process itself; accordingly, given these experimental conditions, there appears to be no need for additional verbal guidance user profiles at this point in the navigational aid design process.

CONCLUSION, OUTCOMES, AND PERSPECTIVES

The combination of the user- and activity-centered approaches in a structured methodology based on Kovacs et al. (2004) allow us to provide considerably improved functional specifications for designing rules for producing verbal instructions and information for blind pedestrians. These rules were formalized and are presented in Gaunet and Briffault (2005) and Gaunet (2006).

Complementary studies are now needed to refine the specifications according to the peculiarities of other visually impaired subpopulations using a "design for all"[12] approach, to identify in greater depth the effects of the use of guide dogs or canes, to design guidance rules for more complex urban environments, and to test the gradual involvement of the interface in the way-finding activity. The interface could, for instance, include alarms in case the pedestrian gets out of the planned path, or in case the localization module has lost the GPS signal[13] (Gaunet & Briffault, 2005), a situation that may occur in urban areas, and some automatic configurable modules that would consider the expertise level of the user (e.g., in Gaunet and Briffault, 2005, we had addressed the issue of wordiness of route description according to user profiles).

The proposed approach can be used for the design and the evaluation of many kind of user interfaces for mobile technology. Actually, mobile use cases that are currently developed consist in new activities to be supported by new services (e.g., tourism and leisure provision of information, eLearning, eWorking, telecare…). They require adequate content models, user and context aware modules, and adequate multimodal interactive devices. Accordingly, appropriate information and interfaces have to be carefully modeled; this is especially true for users with motor, perceptual, and cognitive disabilities and for aging people. The approach presented here shows the decisive importance of involving end-users, human factors researchers and engineers together, from the beginning of the design process to the delivery of the final product. The proposed methodology provides a structured method of gathering professionals around a concrete object (i.e., the activity to be supported by the future system). It thus offers an "interlinguae" that can contribute to the success of the project, provided that some prerequisites, such as awareness of mutual constraints, abilities and knowledge, willingness to work together in a multidisciplinary environment, and creativity are fulfilled.

Overall, our results enlighten the debate opposing two politics for improving way finding: equipping urban areas (talking traffic lights and buildings with RFID[14] tags, tactile ground surfaces…) or equipping visually impaired people. They clearly show the possibility of efficiently guiding them through mobile technology. In addition, such technology can be used for serendipitous exploration of an unfamiliar environment. Coupling currently available devices, such as geographic localization, high precision 3-D-GIS, electronic compasses, and telemetric canes, will allow accurate perception and description of the environment in addition to more accurate and better-oriented locomotion toward a goal. Contrary to locally equipping urban space, both limited in coverage and expensive, mobile technology offers a unique opportunity to visually impaired people for both social and e-inclusion.

REFERENCES

Ashmead, D., & Wall, R. S. (1999). Auditory perception of walls via spectral variations in the ambient sound field. *Journal of Rehabilitation Research and Development*, *36*(4), 313-322.

Ashmead, D. H., Wall, R. S., Eaton, S. B., Ebinger, K. A., SnookHill, M. M., & Guth, D. A. (1998). Echolocation reconsidered: Using spatial variations in the ambient sound field to guide locomotion. *Journal of Visual Impairment and Blindness*, *92*(9), 615-632.

Bentzen, B. L. (1997a). Environmental accessibility. In B. Blasch, W. Wiener, & R. Welch (Eds.), *Foundations of orientation and mobility* (pp. 317-356). New York: American Foundation for the Blind.

Bentzen, B. L. (1997b). Orientation aids. In B. Blasch, W. Wiener, & R. Welch (Eds.), *Foundations of orientation and mobility* (pp. 284-316). New York: American Foundation for the Blind.

Bisseret, A., Sebillotte, S., & Falzon, P., (Eds.). (1999). *Techniques pratiques pour l'étude d'activités expertes*. Toulouse: Octarès Editions.

Brusnighan, D. A., Strauss, M. G., Floyd, J. M., & Wheeler, B. C. (1989). Orientation aid implementing the Global Positioning System. In S. Buus (Ed.), *Proceedings of the Fifteenth Annual Northeast Bioengineering Conference* (pp. 33-34). Boston: IEEE.

Caplan, S. (1990). Using focus group methodology for ergonomic design. *Ergonomics*, *33*, 527-533.

Carroll, J. M., Koenemann-Elliveau, J., Rosson, M. B., & Singley, M. K. (1993). Critical incidents and critical themes in empirical usability evaluation. In J. L. Alty, D. Diaper, & S. Guest (Eds.), *People and computers*, vol. 3 (pp. 279-292). Cambridge: Cambridge University Press.

Carroll, J. M., & Rosson, M. B. (2002). *Usability engineering, scenario-based development of human-computer interaction*. San Diego: Academic Press.

Clark, L. (1991). The use of scenarios by user interface designers. In D. Diaper, & N. Hammond (Eds.). *People and computers*, vol. 1 (pp. 103-115). Cambridge: Cambridge University Press.

Collins, C. C. (1985). On mobility aids for the blind. In D. H. Warren, & E. R. Strelow (Eds.), *Electronic spatial sensing for the blind* (pp. 35-64). Dordrecht: Martinus Nijhoff.

Davis, S. F. (Ed.). (2003). *Handbook of research methods in experimental psychology*. Malden, MA: Blackwell.

Fetterman, D.M. (1998). *Ethnography: Step by step* (2nd ed.). Newbury Park: Sage Publications.

Flanagan, J., (1954). The critical incident technique. *Psychological Bulletin*, *51*, 28-35.

Foulke, E. (1982). Perception, cognition, and mobility of blind pedestrians. In M. Potegal (Ed.), *Spatial orientation: Development and physiological foundations* (pp. 55-76). New York: Academic Press.

Freedom Scientific Web site. Retrieved January, 24, 2007, from http://www.freedomscientific.com/

Fruchterman, J. (1996, March). *Talking maps and GPS systems*. Paper presented at The Rank Prize Funds Symposium on Technology to Assist the Blind and Visually Impaired. Grasmere, Cumbria, England.

Gaunet, F. (2006). Verbal guidance rules for a localized wayfinding aid intended for blind pedestrians in urban areas. *Universal Access in the Information Society*, *4*(4), 338-353.

Gaunet, F., & Briffault, X. (2001). Méthodologie de spécification pour un système informatique de guidage verbal: Le déplacement des déficients visuels en ville. In *Proceedings of EPIQUE 2001, Journées d'Etudes en Psychologie Ergonomique* (pp. 133-143). Poitiers: Société Française de Psychologie.

Gaunet, F., & Briffault, X. (2002). Navigational aid for blind pedestrians: Towards a localized verbal guidance system. *Modeling, Measurement and Control*, *63*(4C), 139-150.

Gaunet, F., & Briffault, X. (2005). Exploring the functional specifications of a localized wayfinding verbal aid for blind pedestrians: Simple and structured urban areas. *Human Computer Interaction*, *20*(3), 267-314.

Golledge, R. G. (1999). Human wayfinding and cognitive maps. In R. G. Golledge (Ed.), *Wayfinding behavior: Cognitive mapping and other spatial processes* (pp. 5-45). Baltimore: Johns Hopkins Press.

Golledge, R., Klatzky, R., Loomis, J., Speigle, J., & Tietz, J. (1998). A geographical information system for a GPS based personal guidance system. *International Journal of Geographic Information System*, *12*(7), 727-749.

Guth, D. A., & Rieser, J. J. (1997). Perception and the control of locomotion by blind and visually impaired pedestrians. In B. Blasch, W. Wiener, & R. Welch R (Eds.), *Foundations of orientation and mobility* (pp. 507-521). New-York: American Foundation for the Blind.

Hatwell, Y., Streri, A., & Gentaz (Eds.) (2003). *Touching for knowing.* Amsterdam: Johns Benjamins Publishing Company.

Helal, A., Moore, S., & Ramachandran, B. (2001). *Drishti: An integrated navigation system for visually impaired and disabled*. Paper presented at the 5th International Symposium on Wearable Computer, Zurich, Switzerland.

Hill, E., & Ponder, P. (1976). *Orientation and mobility techniques: A guide for the practitioner.* New York: American Foundation for the Blind.

Holland, S. Morse, D.R., & Gedenryd, H. (2002). Audio GPS: Spatial audio navigation in a minimal attention interface. *Personal and Ubiquitous Computing*, *6*, 253-259.

Hughes, J.-F. (1989). *Déficience visuelle et urbanisme, l'accessibilité de la ville aux aveugles et mal-voyants.* Paris: Jacques Lanore.

HumanWare Web site. Retrieved January, 24, 2007, from http://www.humanware.ca/web/en/p_DA_Trekker.asp

Jansson, G. (2000). Spatial orientation and mobility of people with visual impairment. In B. Silverstone, M. A. Lang, B. Rosenthal, & E. E. Faye (Eds.), *The lighthouse handbook on visual impairment and rehabilitation* (pp. 379-397). New York: The Lighthouse and Oxford University Press.

Jordan, P. W., Thomas, B., Weerdmeester, B. A., & McClelland, I. L. (Eds.) (1996). *Usability evaluation in industry.* London, UK: Taylor & Francis.

Kovacs, B., Gaunet, F., & Briffault, X. (2004). *Techniques d'analyse de l'activité pour l'interaction homme-machine*. London: Hermes Science Publisher.

LaPierre, C. (1998). *Personal navigation system for the visually impaired*. Master's thesis, Department of Electronics, Carleton University.

Long, R. G. (2001). Spatial orientation and individuals with visual impairments. In D. A. Olson, & F. DeRuyter (Eds.), *Clinician's guide to assistive technology.* Saint Louis: Mosby.

Loomis, J. M. (2005). *User-centered design of assistive technology*? Paper presented at the World Congress on Blind Wayfinding Technology, Baltimore.

Loomis, J. M., Golledge, R. G., & Klatzky, R. L. (2001). Navigating without vision: Basic and applied research. *Optometry and Vision Science, 78*, 282-289.

Macaulay, L. A. (1996). *Requirements engineering.* Berlin: Springer-Verlag Series on Applied Computing.

Mace, R. (1998). Universal design in housing. *Assistive Technology*, *10*(1), 21-28.

Makino, H., Ishii, I., & Nakashizuka, M. (1996). Development of navigation system for the blind using GPS and mobile phone connection. In *Proceedings of the 18th Annual Meeting of the IEEE EMBS*, Amsterdam.

Maulsby, D., Greenberg, S., & Mander, R. (1993). Prototyping an intelligent agent through Wizard of Oz. In S. Ashlund, K. Mullet, A. Henderson, E. Hollnage, & T. White (Eds.), *Human Factors in Computing Systems—INTERCHI' 93* (pp. 277-284). ACM & IFIP TC.13. New York: Addison-Wesley. Amsterdam: IOS Press.

May, M. (2002). Accessible GPS for the blind: What are the current and future frontiers? In *Proceedings of the Center On Disabilities Technology and Persons With Disabilities Conference*, South Africa, PS40–41.

Millar, S. (1994). *Understanding and representing space: Theory and evidence from studies with blind and sighted children*. Oxford: Clarendon.

Nielsen, J. (1993a). *Usability engineering.* Cambridge: Academic Press.

Nielsen, J. (1993b). Evaluating the thinking-aloud technique for use by computer scientists. *Advances in human-computer interaction, 3*, 69-82.

Petrie, H., Johnson, V. Strothotte, T. Raab, A., Fritz, S., & Michel, R. (1996). MoBIC: Designing a travel aid for blind and elderly people. *Journal of Navigation, 49*, 45-52.

Preece, J., Rogers, Y., Sharp, H., Benyon, D., Holland, S., & Carey, T. (1994). *Human computer interaction*. Reading, MA: Addison-Wesley.

Rieser, J. J. (1999). Dynamic spatial orientation and the coupling of representation and action. In R. G. Golledge (Ed.), *Wayfinding behavior: Cognitive mapping and other spatial processes* (pp. 168-190). Baltimore: Johns Hopkins Press.

Rubin, J. (1994). *Handbook of usability testing.* New York: Wiley & Sons.

Rubin, H., & Rubin, I. (1995). *Qualitative interviewing: The art of hearing data*. San Diego: Sage Publications.

Sendero Web site. Retrieved January 24, 2007, from http://www.senderogroup.com/

Sinclair, M. A. (1975). Questionnaire design. *Applied Ergonomics, 6*, 73-80.

SWAN Web site. Retrieved January 24, 2007, from, http://sonify.psych.gatech.edu/research/bonephones/index.html

Thinus-Blanc C., & Gaunet, F. (1997). Space representations in the blind: Vision as a spatial sense? *Psychological Bulletin*, *121*, 20-42.

Walker, B. N., & Lindsay, J. (2006). Navigation performance with a virtual auditory display: Effects of beacon sound, capture radius, and practice. *Human Factors, 48*(2), 265-278.

Wayfinder Web site. Retrieved January 24, 2007, from http://www.wayfinder.com/

Wiener, W. R., & Lawson, G. (1997). Audition for the traveler who is visually impaired. In B. Blasch, W. Wiener, & R. Welch. (Eds.), *Foundations of orientation and mobility* (pp. 104-169). New York: American Foundation for the Blind.

Wiener, W., R., Lawson, G., Naghshineh, K., Brown, J., Bischoff, A., & Toth, A. (1997). The use of traffic sounds to make street crossings by persons who are visually impaired. *Journal of Visual Impairments and Blindness, 91*, 435-445.

KEY TERMS

Activity-Centered Approach: A methodology for the design of systems and their interfaces, based on the actual activity of the end-users in the targeted context and in terms of the user profile, the environment and the goal-directed activities performed. The latter are in-dissociable and form the core of the design process.

Navigational Aid: Any device that supports way finding in an unfamiliar environment

Oriented Locomotion: Locomotion is the movement of one's body around an environment, coordinated specifically to the local or proximal surrounds, the environment that is directly accessible to our sensory and motor systems at (a) given moment(s) (or, at most, within a few moments). It solves behavioral problems such as identifying surfaces to stand on, avoiding obstacles and barriers, directing our movement toward perceptible landmarks, and going through openings without bumping into the sides. During oriented locomotion, people attend to their surrounds, to landmarks, and to their own movement, to reach a perceivable goal.

Rules for Producing Verbal Instructions: Formal production rules that specify the linguistic forms to be provided to pedestrians for way finding and the moment to provide them.

Techniques for the Analysis of the Activity: These are field study approaches that allow gathering information on the needs and wants of end-users, as well as the subject's organization of the activity (the mental operations performed, the informational items to which these operations apply and the resulting physical actions performed).

User-Centered Approach: A methodology for the design of systems and their interfaces, based on the generic characteristics of the end-users (often gathered out-of-context) used to decide on which model the design of the device is to be based.

Visually Impaired Pedestrians: Persons suffering from total blindness, central or peripheral visual deficiency… that will differentially impact their abilities to perform oriented locomotion tasks.

Wayfinding: In contrast to locomotion, way finding is the goal-directed and planned movement of one's body around an environment in an efficient way. Way finding requires a place goal, a destination we wish to reach. This destination is not in the local surrounds. Way finding is coordinated distally, beyond the local surrounds directly accessible to our sensory and motor systems at a given moment. Memory traces of the surroundings, internally or externally stored in artifacts such as maps, play a critical role in way finding. When we way find, we solve behavioral problems involving explicit planning and decision-making, problems such as choosing routes to take, moving toward distal landmarks, creating shortcuts, and scheduling trips and trip sequences.

ENDNOTES

1 Endnotes are used to present the way we have used the techniques presented, as well as the corresponding references. Readers already familiar with these techniques may skip them.

2 *Interviews* (Macaulay, 1996; Rubin and Rubin, 1995) and *Brainstorming* (Caplan, 1990; Macaulay, 1996) techniques, consist of face-to-face or discussion groups led by a moderator. *Interviewing* is performed in order to gain information about the needs or requirements in relation to the new system. In *nonstructured interviews*, the moderator keeps the discussion focussed on the topic addressed.
Brainstorming brings together a cross-section of stakeholders in a discussion group format. This method is useful for eliciting requirements and can help to identify the issues needing to be tackled. The general idea is that each participant stimulates ideas in the other participants, and that through the discussion process, a collective view is established which is greater than the individual parts.

3 *Questionnaires* allow collecting opinions on both needs and usability problems (Nielsen, 1993a; Sinclair, 1975). They consist of lists of open or closed questions on a specific topic. They primarily address representations of the activity not the activity itself. They can inform on either general or precise information, but in the latter case, the domain must be previously mastered to get reliable and useful answers.

4 The *verbalization* technique, also known as the *thinking aloud* technique (Nielsen, 1993ab), consists of asking participants to verbalize their activities in details, either during the task itself or after the task has been performed. This tool allows to evidence the processes (mental operations) as well as informational items to which they apply, at least those that can be verbalized.

5 The *scenario* technique (Carroll and Rosson, 2002; Clark, 1991) primary aim is to provide examples of projected use of the tool in development to understand and clarify user requirements, and to provide a basis for later usability testing. Errors, verbalisations and behavioral cues are recorded to understand both successes and failures. It thus helps to identify usability targets and likely task completion times before a commitment to code writing is made. Scenarios should be based on the most important tasks from the context-of-use information. User goals are decomposed into the operations needed to achieve them.

6 The *observation* technique (Jordan, Thomas, Weerdmeester and McClelland, 1996; Nielsen, 1993a) consists of formally describing the sequence of behavioural units performed by the participant. A prerequisite is to define the appropriate behavioural units to be recorded, what will be done according to the granularity of the task supported by the future interface. Knowing the behavioural sequence of a task, whether already supported by an interface or not, gives insight into when, what and how to provide information to the user.

7 The *information-on-demand* technique (Bisseret, Sebillote and Falzon, 1999) consists in providing a participant the information he/she requires to perform the steps of his/her activity, when and only when, he needs it and asks for it. The relevance of the information provided can thus be directly evaluated.

8 In the *wizard of Oz* technique (Maulsby, Greenberg and Mander, 1993; Nielsen, 1993a), an experimenter (the "wizard"), in a laboratory or real setting, simulates the behavior of a theoretical intelligent computer application (often by staying out of the user's perception and intercepting all communications between participant and system). The approach is particularly suited to explore design possibilities which are demanding to implement such as intelligent interfaces possibly featuring agents or advisors, and/or natural language processing.

9 *Experimentation* is a classical technique (Davis, 2003; Rubin, 1994). It consists in testing the effects of one or more controlled factors of the task declined in a few modalities, for example, the effects of the color of the buttons, of different types of interactive joysticks. It is usually used to address specific questions.

10 *The impromptu recall* technique (Bisseret et al., 1999) consists of interrupting the participant when he/she is performing a task and asking him/her what he/she was doing or thinking about.

11 The *critical incidents* technique (Carroll, Koenemann-Elliveau, Rosson and Singley, 1993; Flanagan, 1954) is based on recording critical incidents, events that represent significant failures of a design. Verbal reports of the incident are analysed and categorized to determine the frequency of different incident categories. This enables design deficiencies to be identified and avoided, as well as identification of actions to be cautiously supported.

12 "Design for all" is an approach to the design of products and environments to be usable by all people, to the greatest extent possible, without the need for adaptation or specialized design (Mace, 1998).

13 Which should occur less and less with forthcoming advances in localization technology such as the European Satellite navigation system Galileo.

14 Radio Frequency IDentification

Chapter XLII
Trends in Adaptive Interface Design for Smart Wheelchairs

Julio Abascal
University of the Basque Country-Euskal Herriko Unibertsitatea, Spain

Borja Bonail
University of the Basque Country-Euskal Herriko Unibertsitatea, Spain

Daniel Cagigas
Universidad de Sevilla, Spain

Nestor Garay
University of the Basque Country-Euskal Herriko Unibertsitatea, Spain

Luis Gardeazabal
University of the Basque Country-Euskal Herriko Unibertsitatea, Spain

ABSTRACT

This chapter introduces the main trends in the design of interfaces for smart wheelchairs. It stresses the importance of taking into account their similarity with autonomous mobile robots and the restrictions posed by the special characteristics of the users (people with severe motor and speech restrictions) and the task (indoor assisted navigation). With this background, the main features of the user-wheelchair interface are discussed, justifying the need for the adoption of an adaptive approach. The TetraNauta smart wheelchair interface design is used to illustrate the design of user, context, and task models. In addition, it describes some mechanisms to enhance the communication rate when a single-switch-scanning input device is used (scanning rate dynamic adaptation and word prediction). The chapter finishes with a discussion of the influence of new navigation models, such as the behaviour-based one, in the design of the user interface.

INTRODUCTION

From the technological point of view, smart wheelchairs are very similar to autonomous vehicles or mobile robots. They have actuators (electric motors), sensors (mainly for localise and distance measuring), and a real-time controller that performs the mapping, planning, and driving tasks. Similarly, to other autonomous vehicles, they are designed to transport individuals, but smart wheelchairs present some interesting differences: due to the special features of the user, they require the necessary interaction between the system and the person to be performed with the minimum physical and cognitive effort.

The intense relationship between the user and the wheelchair raises interesting challenges to the designers of the user interface. The user must be in control, while leaving many navigational tasks to the system. The dialogue must be designed in a way that allows the user to give commands quickly enough to efficiently interact with a mobile system, with the minimum physical and cognitive effort.

The following sections present the technological context for which smart wheelchairs are designed, and the constraints and challenges that the wheelchair-user interface design must face, illustrated by details of the design of the TetraNauta smart wheelchair interface (Abascal, Cagigas, Garay, & Gardeazabal, 1999).

WHAT IS A SMART WHEELCHAIR?

Electric-powered wheelchairs are prescribed to people with lower-limb motor disabilities who have difficulties using manual wheelchairs due to upper-limb motor disabilities, such as arm, hand, shoulder, or more general movement restrictions. The user of an electric wheelchair typically controls its speed and direction by operating a joystick on a controller. If the user lacks coordination or strength in the hands or fingers, a number of other input devices can be used, such as chin controls or puff/suck scanners. However, due to severe physical or cognitive restrictions, several users find it extremely difficult to control powered wheelchairs by means of conventional methods, especially when they must perform complex steering or manoeuvring tasks. Smart wheelchairs are mainly devoted to avoiding these problems (Fehr, Langbein, & Skaar, 2000).

Therefore, smart wheelchairs are designed to improve the autonomous mobility of users with severe motor impairments (e.g., quadriplegia or multiple sclerosis) who experience difficulties in driving traditional electric wheelchairs. Individuals in this population can also experience low vision, visual field reduction, spasticity, tremors, or cognitive deficits.

Even though smart wheelchairs have occasionally been used for training and evaluation of traditional and electrically powered wheelchairs (Simpson, 2005), their main purpose is to assist the user with navigation tasks in real time. Since the remaining capabilities of the users are very diverse, and they can change with time due to fatigue, disease evolution, changes in motivation, and so forth, the type of assistance required is also very diverse. Some people may only require help to traverse narrow places, such as doors, or to get out from difficult places, such as small toilets, while others may require complete navigation help, including location and mapping, path planning, and accurate driving to a selected destination. Therefore, smart wheelchairs should be able to perform navigational tasks autonomously, though always taking into account the user's needs, capabilities and desires.

Smart Wheelchair Structure

A smart wheelchair typically consists of either a standard powered wheelchair to which a computer and a collection of sensors have been added, or a mobile robot platform to which a seat has been attached (Simpson, 2005). The majority of smart wheelchairs developed nowadays are based on commercially available powered wheelchairs. Some models are even designed as independent units that can be attached to and removed from the commercial wheelchair. Several relevant examples of smart wheelchairs can be found in the literature, such as NavChair (Levine, Bell, Jaros, Simpson, Koren, & Borenstein, 1999), SENARIO (Katevas, Sgouros, Tzafestas, Papakonstantinou, Beattie, Bishop. et al., 1997), Hephaestus (Simpson, Poirot, & Baxter, 2002), TetraNauta (Vicente-Diaz, Amaya-Rodríguez, Díaz-del-Río, Civit-Balcells, & Cagigas, 2002) and SPAM (Simpson, LoPresti, Hayashi, Guo, Ding, & Cooper, 2003).

The references show that smart wheelchairs have been studied and developed since the early 1980s. In order to solve the navigational needs of people with severe disabilities, smart wheelchairs have been equipped with sensors, artificial vision devices, controllers, and human-robot interaction systems, using techniques and devices taken from the mobile robotics field. Most of them have not progressed beyond the prototype stage because the complex extra hardware added makes the resulting wheelchair extremely expensive and frequently difficult to drive.

The high economic cost is the main, but not the only, problem that smart wheelchairs must face. Technical problems, such as the lack of a standard open communication protocol for powered wheelchairs, make it difficult to develop simple interfaces between computer/electronic technology and currently manufactured wheelchairs. Another associated problem is performance evaluation in actual environments with real users. A proper performance evaluation of a smart wheelchair implies long-term studies with several users and wheelchair units. Therefore, these extensive resources imply prohibitive costs that only major wheelchair manufacturers can afford. Despite the problems and disadvantages, smart wheelchair technology is now mature and ready to be applied, especially to some specific fields, such as indoor navigation in structured environments.

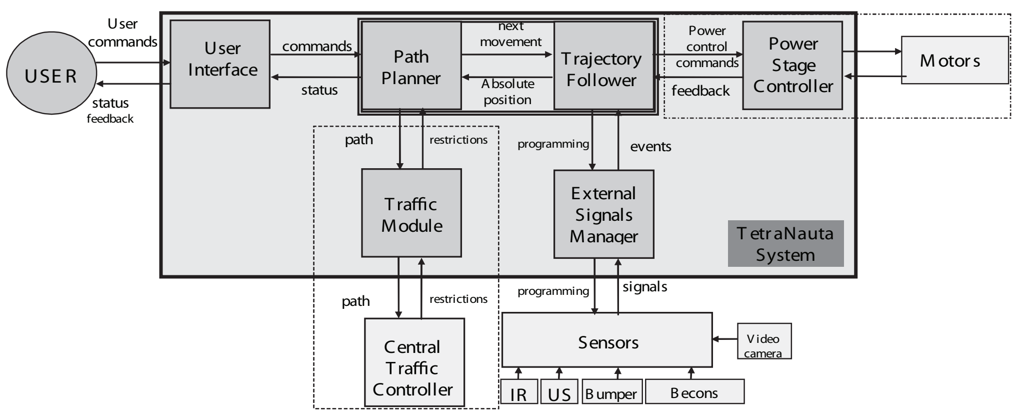
More information about smart wheelchairs can be found in Ding and Cooper (2005), and in the comprehensive literature review published by Simpson (2005). In this chapter, we will refer to the experiments performed with the TetraNauta smart wheelchair (view figures 1 and 2), designed within the successive TetraNauta I, II and III projects by the Universities of Seville and the Basque Country (Vicente-Díaz et al., 2002).

Smart Wheelchair User Interface

Focusing on human-smart wheelchair interaction, it is evident that, despite the wheelchair having autonomous navigation, user involvement is still necessary to maintain high-level task control. A good smart wheelchair, similar to a mobile robot, is able to navigate autonomously. However, contrary to autonomous mobile robots, the smart wheelchair has to perform the navigation taking into account the current user's wishes. To be able to do this, a "shared control" paradigm must be adopted. That is, the navigation must be controlled by both the system and the user in a collaborative fashion.

If the wheelchair has to follow the orders from the user, an efficient system of communication

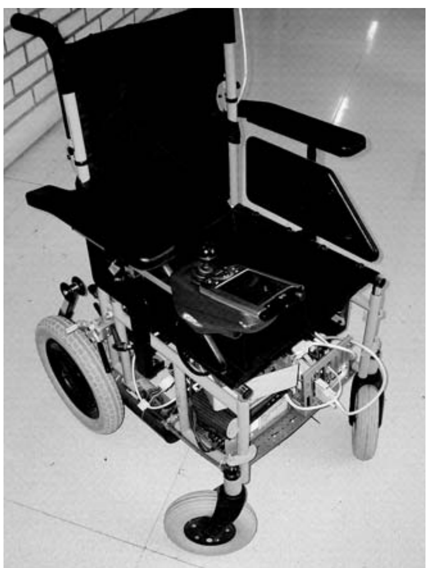
Figure 1. Architecture of the TetraNauta smart wheelchair



between both of them is required. As aforementioned, people with severe motor restrictions usually experience difficulties in using standard input/output devices. In fact, numerous wheelchair users have severe physical restrictions also affecting speech communication. They tend to control the wheelchair by means of alternative devices that have been designed for alternative and augmentative communication[1].

Input devices of this kind, such as one-key-scanning systems, are often too slow to move through the enormous number of different possible orders and destinations, arranged in standard comprehensive menus that can be selected by the user. It is therefore necessary to take advantage from context restrictions in order to reduce the set of choices, consequently facilitating the communication. Intelligent interfaces are good candidates for this task because they are able to use the knowledge of the current conditions (Yanco & Gips, 1998). For instance, the subset of destinations reachable from the present position, user's habits, characteristics of the current activity, and so forth, can be used to assist the user with command selection.

Figure 2. A prototype of the TetraNauta smart wheelchair



In addition, the design of the interface has to take into account the restrictions imposed by the type of system (usually a mobile small device, such as the PDA proposed in Myers, Wobbrock, Yang, Yeung, Nichols, & Miller, 2002), the features of the users (people with severe motor, and frequently, speech restrictions) and the task they are performing (habitually the selection of a destination from a number of choices available in a structured environment).

USER INTERFACE DESIGN CONSTRAINTS

Let us analyse the special features of the users and the restrictions imposed by the physical, social, and economical context that must be taken into account when facing the design of the user interface for a smart wheelchair.

Disability and Rehabilitation

The difficulties that physically disabled people have when handling standard mobile interfaces are usually due to the requirements of coordination and strength in the upper limbs imposed by traditional input systems. Therefore, these users require alternative input devices and procedures adapted to their own features. In addition, it is necessary to automate interaction tasks as much as possible in order to minimise user intervention (Abascal, Cagigas, Garay, & Gardeazabal, 2002). These issues are described in a subsequent section.

In addition to physical characteristics, there are some issues related to user attitudes, interests, cognitive evolution, and so forth, that must be taken into account in order to ensure the appropriateness of the interface. For instance, even if many interactive tasks can be automated, it is necessary to ensure user participation in order to enhance his or her interest, participation, and rehabilitation:

- **Control:** The user must feel that he or she is the one who decides. Otherwise, the device can generate frustration and passivity. That includes ease of switching between automatic/assisted/manual functioning.

- **Rehabilitation:** Numerous people with disabilities are able to enhance their cognitive abilities, personal attitudes, and social integration when they are provided with adequate user interfaces (Butler, 1986). Since these aspects can produce quite rapid changes in the interaction, the interface should be able to adapt to people's physical and cognitive variations.
- **Security:** Several of these systems interact with the environment in various ways that can be dangerous in the case of failure or malfunction. The designer must ensure that the system is safe, reliable, and fault tolerant.

Price

The design of a smart wheelchair is frequently approached as if it was just a mobile robot. That means that sophisticated and expensive sensors are incorporated, and the original elements of commercial electric wheelchairs, such as the power stage or the control bus, are modified. The result of such an approach is usually a high quality and high performance autonomous mobile robot, but one which is frequently too complex to handle and too expensive to be marketed for people with disabilities.

Therefore, nonexpensive solutions are necessary in order to avoid unaffordable systems. This means using less expensive sensors (e.g., infrared and ultrasonic rather than laser sensors, to measure distances) and enhancing the "smartness" of the control program. Nowadays processors are cheap, and the lower quality of the sensors is balanced by a much higher processing capacity.

On the other hand, since most smart wheelchairs are built over commercial electric wheelchairs, it is convenient to avoid large modifications in order to facilitate their potential future manufacture without the need for improbable great investment by the wheelchair industry.

Many designers tend to ignore such socioeconomic issues because they are not purely technical. Nevertheless, they can be crucial for ensuring the usability of the resulting device.

Shared Control

As has been previously mentioned, the goal of smart wheelchairs is to automate, as much as possible, the navigation operation, while leaving in the user's hands all those tasks that he or she is able to perform. This facilitates user rehabilitation by promoting his or her involvement and avoiding passivity. This kind of collaboration for control is called *shared control* in mobile robotics (Yanco & Drury, 2004). In the shared control paradigm, tasks are performed by both the user and the system in perfect harmony. Usually, the user performs the high-level tasks (e.g., planning) and the system the low-level ones, which are usually tedious or require a high degree of accuracy (e.g., driving). This means that the user is the one who makes decisions, except possibly in hazardous situations.

Shared control can be achieved using preset schemes that can be defined for simple cases. For instance, Baker and Yanco (2004) state: "In shared mode, the robot drives itself, avoiding obstacles. The user, however, can influence or decide the robot's travel direction through steering commands." In the case of a smart wheelchair used by a person who only has problems in driving across difficult passages or avoiding obstacles, an alternative scheme can be applied. The user normally drives the wheelchair, and the system is programmed to gain control only when an obstacle or a narrow space is detected. Even in this case user, participation is still necessary. To cross a doorway, for instance, the user drives the wheelchair through it, but the system makes the required corrections in trajectory and speed to avoid collisions. These schemes can be implemented using standard wheelchair input devices, such as joysticks, *mouthsticks*, and similar devices. The use of any kind of feedback information could also be useful for the user.

Shared Control Schemes

The navigation task is usually divided into *global* and *local* navigation subtasks. The global navigation is to determine one's position in absolute or map-referenced terms, and to move to a desired destination point (path planning), while the local navigation is to determine one's position relative to objects in the environment, and to interact with them correctly (driving). It is possible to classify the different shared control schemes, taking into account who is in charge (user, robot, or both) of global and local navigation tasks, respectively (see Table 1).

Each of these pairs has its own navigation features and therefore needs a different user interface, as shown:

- **Global navigation user (GU):** The user is able to plan the desired path either in known or unknown environments. In known environments, the path planning is usually performed before the navigation, and it is based on a structured map of the world. In unknown environments, the path planning is usually performed continuously during the navigation and is biologically inspired, based on well-known concepts (see the section on behaviour-based interaction). **Interface requirements:** Only if the user is going to pass the desired path to the robot will he or she need a suitable input method according to his or her skills, like a touch screen or speech recogniser. If the path is generated during the navigation, the user will need a simple interface to provide the steering commands to the robot (See, for instance, Yanco and Gips, 1998).
- **Global navigation shared (GS):** The user is unable to plan the desired path on his or her own, either in known or unknown environments, but they are able to choose the desired one from a list generated by the robot. **Interface requirements:** The user will need a suitable selection mechanism like a "one-key-scanning" input, for example, if he or she has problems with the standard ones such as touch screens or joysticks.
- **Global navigation robot (GR):** The user is unable to plan the desired path either in known or unknown environments, so the robot generates a suitable path. **Interface requirements:** Only if the user wants to know the planned path will he or she need a suitable output method according to his or her skills, such as a graphic display or text-to-speech system.
- **Local navigation user (LU):** The user is able to drive the wheelchair correctly. **Interface requirements:** This kind of user will use standard drive interfaces like joysticks and similar devices.
- **Local navigation shared (LS):** The user can drive easily in open environments, and difficult and hazardous situations are performed by/with the robot. **Interface requirements:** In order to know what the robot is doing when it takes control, some kind of feedback to the user will be very useful.
- **Local navigation robot (LR):** The user is unable to drive, but may provide the steering commands according to the planned

Table 1. Shared control schemes

		Local Navigation (drive)		
		User	**Shared**	**Robot**
Global Navigation (plan)	**User**	Teleoperation	GU-LS	GU-LR
	Shared	GS-LU	GS-LS	GS-LR
	Robot	GR-LU	GR-LS	Full autonomy

path. **Interface requirements:** This kind of user will use standard drive interfaces like joysticks and similar.

Adjustable Autonomy

Nevertheless, when the user has more significant restrictions, or when he or she experiences changes in his or her capabilities, there is a need for more complex control sharing, where the degree of autonomy is negotiated continuously. Baker and Yanco (2004) call it *adjustable autonomy* to a sliding balance between robot autonomy and mixed initiative. In this case, the system needs to know more about the current status of the user in order to adapt its operation; what the user wants to do and what he or she can do currently. For example, when the system detects that it is approaching an open door, it should know whether the user wants to cross it or not and, if so, the amount of help he or she needs in order to traverse it. In order to avoid mode confusion, a signalling mechanism is needed in the user interface. In other words, when the system is changing automatically between different modes of shared control, the user must be informed about the current mode of the system to avoid misunderstanding situations.

Complex shared control requires adaptive behaviour from the system. That is, a model to dynamically combine user wishes and system decisions, provided that the intelligent interface is able to interpret the user's intentions. The latter is the function of the user-wheelchair intelligent interface.

Human-Robot Interface Design Guidelines

In the same way that smart wheelchair technology is mainly based on developments in mobile robotics, user interface design for smart wheelchairs is able to draw upon various developments that have taken place in the human-robot interface field. For instance, Drury et al. (Drury, Hestand, Yanco,& Scholtz, 2004) provide excellent human-robot interface design guidelines.

- **Enhance awareness:** Provide a map of where the robot has been. Provide more spatial information about the robot in the environment to make operators more aware of their robots' immediate surroundings.
- **Lower cognitive load:** Provide fused sensor information to avoid making the user fuse the data mentally.
- **Provide help in choosing robot modality:** Provide the operator with assistance in determining the most appropriate level of robotic autonomy at any given time.

Some of these guidelines are particularly useful for user-smart wheelchair interface design. However, they must be applied taking into account that the mental model of the user is not a reflection of the environment model of the robot. For instance, people may not have a structured map of the building in their minds. They may need location information presented as relative to recognizable environmental elements.

ADAPTABILITY

As has already been discussed, since smart wheelchair users are usually people with severe motor restrictions; they need user interfaces that are able to interpret their commands with the minimum effort. Intelligent user interfaces are able to adapt their operation to the current situation, based on the actual values of diverse parameters related to the user and the context, and avoiding the need for verbose interaction.

The system uses for adaptation a model composed of parameters that are *observable* and *relevant* to the interaction. These parameters may take a set of values that are defined by the designer. A specific combination of these values determines the current status of the modelled entity. In addition, a set of rules defines the behaviour of the system for each status, allowing dynamic adaptation to the temporal changes.

The adaptability of the smart wheelchair user interface is based on relevant information, such as the characteristics of the user, the context, and the task that the user is performing at this precise time in that context. Let us briefly describe three models used for the TetraNauta Smart wheelchair interface.

User Modelling

A simple user model based on traditional user modelling techniques is used for the TetraNauta interface. After a study of the users' features[2], five relevant and observable parameters were selected: *reaction* capacity (slow, medium, fast), upper limbs movement *precision* (low, high), *speech* intelligibility (dysarthric, intelligible), number of *keys* that they can type (1, 2, 9), and *orientation* capacity (low, high). In this way, each user is defined by a 5-tupla of parameters:

user (*reaction, precision, speech, keys, orientation*)

For instance, the expression:

user (slow, low, dysarthric, 1, low)

defines a type of user with a very severe motor and speech disability.

Diverse user profiles and user stereotypes were defined in function of the possible meaningful combinations of parameter values. For instance, the expression:

stereotype (intelligible)

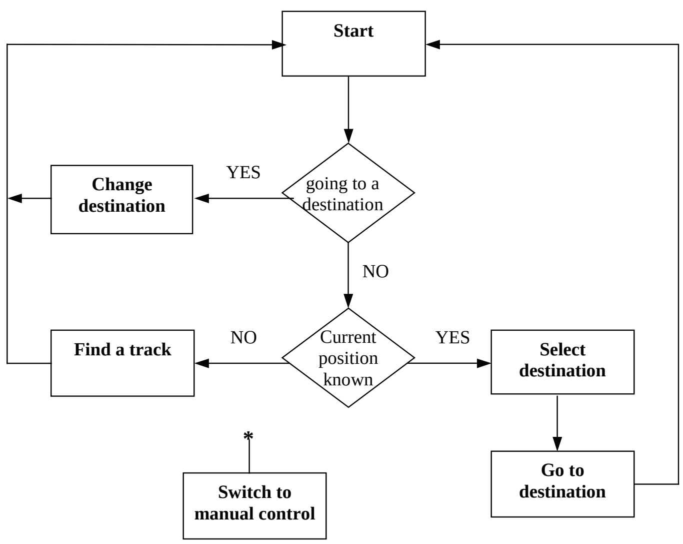
groups all the users that are able to speak, forgetting all the other parameters.

Two of these parameters, *reaction* time and *precision*, can change rapidly due to variations in fatigue or motivation, while *orientation* varies smoothly with the learning process. The system estimates the current value of these parameters in order to characterise the user and includes him or her in the relevant stereotypes. It uses a small inference engine to make assumptions about the user and to dynamically configure the interface, including deciding what kind of tasks can be left to the user and what must be done automatically by the system.

Task Modelling

The set of tasks that can be performed within TetraNauta include *automatic driving, manual driving, remote voice communication, remote/local text communication, environmental control,* and *access to the Web.* The system is able to identify the active task and provide the most suitable user commands for it.

Figure 3. Simplified task diagram for TetraNauta smart wheelchair



Task modelling also allows the simplification of the interface. For a particular status, only the possible actions are offered to the user. For instance, when the wheelchair is automatically driving the user to a destination, only "change destination" and "switch to manual control" commands are offered, minimising the selection effort. A schema of the simplified task diagram is shown in Figure 3.

TetraNauta task modelling is really simple and does not need a reasoning mechanism. An estate machine is enough to model the possible command in function of the current task.

Context Modelling: Maps and Paths

Since navigation is the main objective of a smart wheelchair, in order to improve the user's mobility limitations, let us describe in more detail the mapping of the spatial model held by the system within the user interface.

Internal Structures for Navigation

Smart wheelchair and mobile robot navigation tasks are almost equivalent. They have to face the same navigation problems and use the same technology. In fact, navigation is a subproblem of a more general robotics problem called robot motion planning (Latombe, 1990).

The key difference between robotic and human navigation is the quantum difference in perceptual capabilities. Current robots, while being able to detect stationary obstacles before they run into them, have very limited perceptual and decisional capabilities (Henlich, 1997). Human involvement is even more necessary in a smart wheelchair navigation task in order to ensure user safety.

Navigation is traditionally divided into subtasks: Environment perception/modelling and positioning; path planning between a starting position and a goal position; and driving/guiding through the previously planned path.

For the first and second subtasks, smart wheelchairs use an internal map of the environment in which they have to navigate. Maps can be classified into topological or metric. They can be constructed online by the robot, using learning methods, or loaded off-line as with car navigation systems. Topological maps are represented by graphs and specify connections between free spaces. Metric maps (also called cell maps) divide the environment into cells. These cells are classified into free space or obstacle.

Positioning can be absolute or relative. In the first case, the robot's position on the map is tracked each time the navigation system needs it. Such methods include the use of, for example, GPS or beacon systems. In the second case, its position is calculated taking into account the last position recorded and the subsequent displacement (i.e., dead reckoning). Usually, robot navigation systems combine both techniques to improve accuracy.

Path planning is associated with, or influenced by, the type of map selected. Thus, path planning with graphs is performed using graph search algorithms (such as the classic A* and its variations). Graph search is also possible with cell maps if a connectivity graph is extracted. Additionally, cell maps allow other path planning techniques, such us potential fields, wavefront expansions, or genetic algorithms.

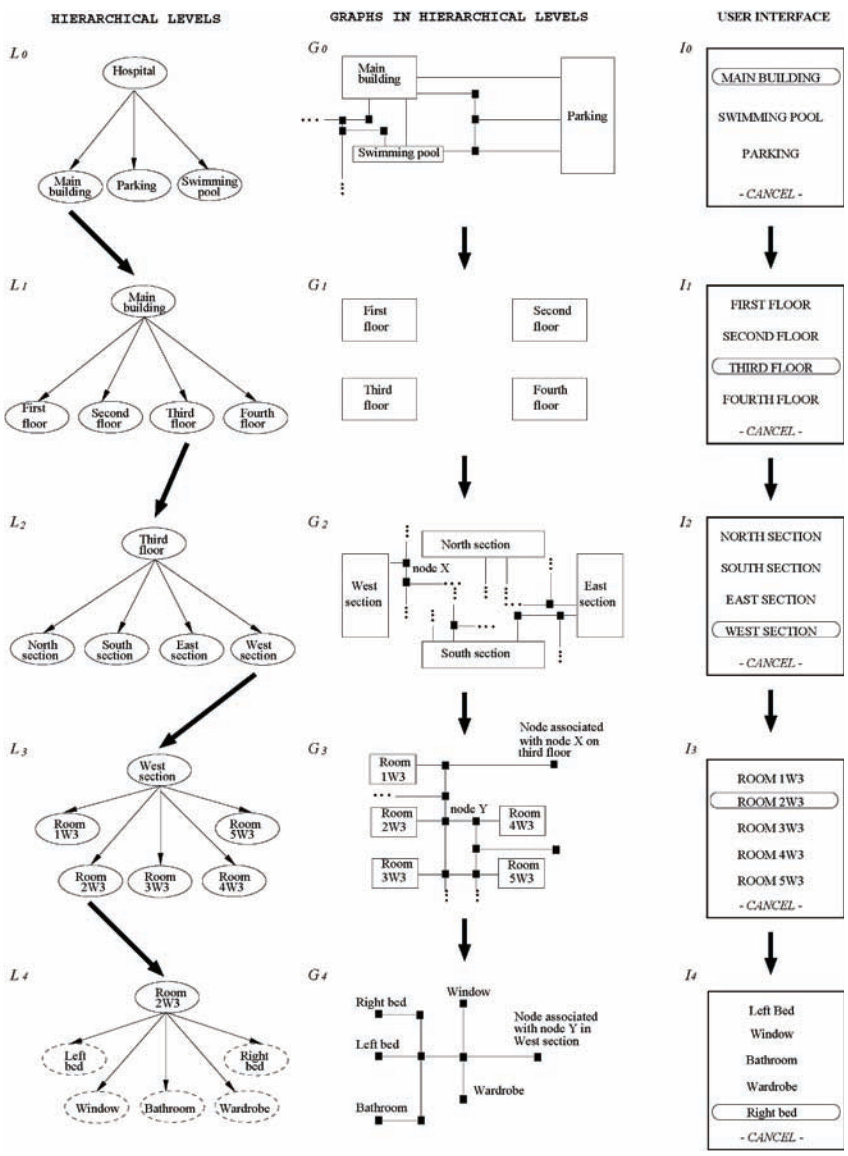
Several techniques can be implemented to drive a smart wheelchair through a path. The most simple guidance methods use sonar to follow walls (and door passages), infrared sensors to follow stripes on the floor, or bump sensors to avoid obstacles, for instance. Other more sophisticated guidance systems use gyroscopes or compasses for orientation, laser range finders to scan obstacles in a 180º two-dimensional plane, or artificial vision to identify special marks or beacons. Due to the lack of accuracy and the cost of some sensors, it is common to use high-level techniques that improve the perception of the environment like machine learning or sensor fusion[3].

Sharing Navigational Information with the User

The way in which information about an environment is perceived and arranged by the navigation system plays an important role in smart wheelchair user interface design. Even if GPS systems cannot be used in indoor environments, their interfaces can be considered for the representation of space in smart wheelchair user interfaces. The idea is to reproduce a topologic map of the environment, indicating the evolution of the wheelchair's position, and asking the user to produce commands based on that information. Even if the user is able to recognise the environment on the map, this system does not facilitate the production of commands by the user. A more abstract representation would allow the user to choose easily from a restricted number of choices to generate navigational commands. The following paragraphs describe how the internal map used by the wheelchair can be also used as a spatial model to communicate with the user.

A good map structure and an efficient path planner can facilitate the interaction between the navigation system and the user interface. Hierarchical graphs (H-Graphs) and hierarchical search algorithms are, for example, a valid solution to

Figure 4. An H-Graph representing a hospital map and its interrelation with the information displayed in a smart wheelchair interface (Adapted from Cagigas and Abascal, 2004)



this problem. An H-Graph divides the information into different levels of abstraction that are isolated from one another except for the abstraction links between them.

This can be used to get rid of unnecessary details, when processing the information, by selecting a level of abstraction where the information can be managed without an excessive amount of data (Fernández & González, 2001). On one hand, search algorithms can take advantage of H-Graphs to speed up path generation and provide a rapid system response, even in critical situations and with large graphs of thousand of nodes and arcs (Cagigas, 2005). On the other hand, environment information can be arranged and transmitted by the interface to the smart wheelchair user in a better and more flexible way. Figure 4 shows an example of a basic display interface used to select and explore reachable zones/points by a smart wheelchair in a hospital. The H-Graph represents a hospital map and its interrelation with a basic smart wheelchair display interface used for navigation. The H-Graph has five abstract levels (L0-L4) and five associated graphs (G0-G4). The different display screens (I0-I4) show the possible zones and points reachable by the smart wheelchair navigation system. The display interface can easily be substituted or completed with a voice synthesizer. The context model allows the simplification of the interface dialogue since at a given position. only certain tasks can be performed. Therefore, the choices that the interface offers to the user are limited to the ones possible at the current location. For example, only destinations reachable from the current position (usually ordered by frequency of use) are offered.

System Behaviour

The combination of the three models allows the simplification of the interface dialogue. The system behaviour, **sb**, which is the set of choices that the interface proposes to a specific user performing a particular task, depends on the user features, **u**, the current task, **t**, and the context in which this task is developed, **c**:

sb = f (u ,t , c)

For instance, the knowledge that the system has about a person (*user*$_i$) trying to select a destination (*task*$_j$) from the current position (*context*$_k$) allows the display of the points reachable from the current position, ordered by user preference or frequency (as described by *system_behaviour*$_h$), graphically displayed in Figure 5.

$sb_h = f(u_i, t_j, c_k)$

In this way, the system provides the user with an adaptive interface to communicate with the wheelchair, displaying the options most likely to be desired by the user performing a specific task at a specific location. If the desired action is one of those displayed by the interface, the effort that the user must make to select it is minimal. In addition, mechanisms are provided to allow selections not predicted by the system.

Using this mechanism, when the user is, for instance, described by:

User_X (slow, low, dysarthric, 1, low)

Figure 5. Three-dimensional space of system behaviour in function of the user, the task, and the context

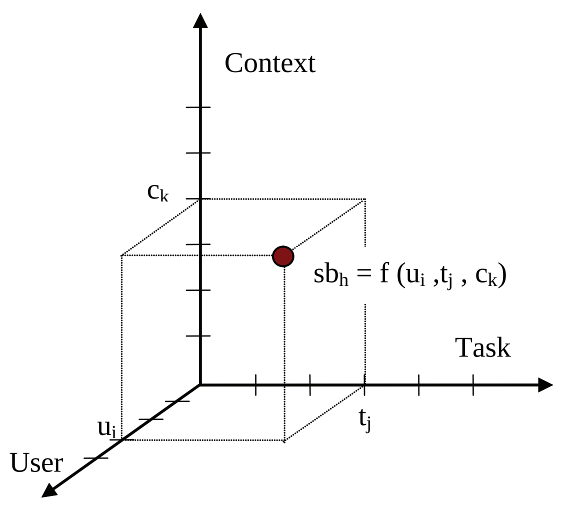


Table 2. A selection matrix, to select a destination from room 2W3

Destination:							
Room 5W3	STOP	START	TO	Matrix2	Matrix3	?	!
TV room	Blank	E	O	R	U	Q	G
Lift	A	S	L	T	B	J	X
Gymnasium	N	D	C	Y	F	W	K
Tavern	I	M	H	Z	P	V	.

the task is:

Task (change_destination)

and the context is:

Context (room_2W3)

the system will adapt the interface to a system behaviour that includes the following characteristics:

- A single-switch-scanning input, with slow scanning period, not using maps for feedback messages.
- The first choices in the matrix of options are the most frequently accessed places from room 2W3 (See table 2).

THE PHYSICAL INTERFACE

In previous sections, designing the dialogue in such a way as to minimise the quantity of information that must be produced by the user was discussed. With the same objective, that is to minimise the effort that the user has to make to control the wheelchair, it is convenient to provide input devices adapted to the user's remaining capabilities in order to take advantage from them to communicate the necessary information.

Input Devices

One of the key points for wheelchair control is the input device used. Since smart wheelchair users usually have severe physical restrictions, the joysticks traditionally used to drive electrically powered wheelchairs are frequently insufficient. The smart wheelchair offers the possibility of performing a number of tasks on behalf of the user. To this end, it is necessary to provide a more complex interface where the user can give his or her commands. Voice communication appears to be the best solution because of the high semantic power of spoken language and the low effort required to produce short orders. The problem is that only a small fraction of potential smart wheelchair users are able to speak in an intelligible fashion.

Most people with motor and speech restrictions use single-switch-scanning interfaces for communication purposes. Scanning systems arrange the set of options in one or more matrices that are row-column scanned. When the desired item is highlighted, the user can select it through the activation system, usually a switch handled by means of any residual controlled movement (in hands, head, breathing, making sounds, etc.)[4].

Enhancing Communication Speed

A number of users from the *Hospital Nacional de Parapléjicos de Toledo*, in Spain, that took part in the development and evaluation of TetraNauta wheelchair use a single-switch-scanning interface[5] for communication purposes. It is well known that this kind of input system is quite slow[6]; thus, efficient user-system communication requires communication acceleration mechanisms. The TetraNauta interface allows the use of single-switch-scanning input with two acceleration methods: word prediction and scanning time adaptation.

Table 3. A selection matrix, after the user has selected "T" character, with four predictions in the first column

T_							
Toilet	STOP	START	TO	Matrix2	Matrix3	?	!
Tavern	Blank	E	O	R	U	Q	G
Tom's	A	S	L	T	B	J	X
Tobacconist	N	D	C	Y	F	W	K
	I	M	H	Z	P	V	.

Word Prediction

Word prediction is one of the most widely used techniques to enhance the communication rate in augmentative and alternative communication. A predictor is a system which tries to anticipate the next block of characters (letters, syllables, words, sentences, etc.) that the user wants to select, in order to reduce the effort and time for message elaboration. In general, prediction is based on the previously produced information, and it is possible due to the redundancy inherent to natural languages.

In order to reduce the effort required, it is necessary to decrease the number of keystrokes needed for composing a message (Garay & Abascal, 2006).

TetraNauta interface includes a word prediction system to assist the user when he or she has to enter names of places, destinations or similar strings of characters in natural language, taking into account the frequencies of letters and words and their context (Garay, 2001). When the user enters a letter, the system analyses the context and proposes to the user the most probable words starting with this letter (see Table 3). If the user does not select any of the proposals, he or she can enter a new letter and the prediction process is repeated. The system dynamically adjusts the frequencies of used words and includes the new words in the dictionary. Due to the small dictionary used for wheelchair navigation proposals, the dependency on the context, the hit rate of the prediction system is very high. Depending on the size of the context only one or two selections are needed to produce the name of a destination.

Scanning Time Adaptation

The scanning of the matrix containing the options is usually done at a fixed rate. Let **T** be the period of time while an option is selectable, $\mathbf{T_r}$ the average time required by the user to react, $\mathbf{T_s}$ the time (usually constant) to select the item, and $\mathbf{T_e}$ the remaining time. It is evident that:

$$\mathbf{T = T_r + T_s + T_e}$$

T should be as short as possible to optimise the communication rate. When the selected T is too short, the user has not enough time to react and produces frequent mistakes, slowing the process. On the other hand, if T is too long, the unnecessary remaining time T_e is large, also slowing the process. Therefore, the best rate is obtained when the most adequate T for each user is found. Studies conducted with users shown that for each specific user, T changes along short periods of time due to diverse factors: attention, fatigue, mood, interest, and so forth. For this reason, there is not an ideal T for each user, but a range $[T_{min}, T_{max}]$. Therefore, the best communication rate would be obtained if the system could dynamically adapt T to the current reaction capacity of the user.

TetraNauta interface is able to adapt the scanning period T to the current reaction time of the user by means of a fuzzy model that takes into account the current and average selection times (Gardeazabal, 2000). To this end the following parameters are used:

- T_{es}: Estimated T
- T_c: Current T
- T_{ar} : Average T_r
- T_n: New T
- D: Deviation of T

A new scanning rate is obtained from the current rate T_c using this formula:

$$T_n = T_c - f\,D$$

where f [0, 1] is an adjusting factor to smooth the transition to avoid sudden changes and D is:

$$D = T_c - T_{es}$$

and

$$T_{es} = 2\,T_{ar} + K$$

where K is a constant offset [0, 1] value necessary for fine tuning, depending on the specific user. Therefore:

$$D = T_c - (2\,T_{ar} + K)$$
$$T_n = T_c - f\,(T_c - T_{es}) = (1\text{-}f)\,T_c + f\,(2\,T_{ar} + K)$$

In particular, with K = 0, the new T is:

$$T_n = (1\text{-}f)\,T_c + 2\,f\,T_{ar}$$

Figure 6 shows the results of the dynamic adaptation for a specific user, based on his reaction time when selecting both characters and blocks of information.

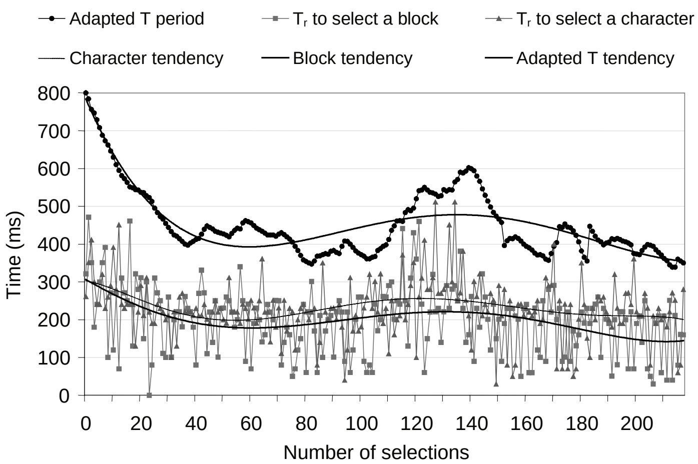
ALTERNATIVE NAVIGATION MODELS AND THEIR INTERFACES

Two Approaches to Navigation and Interface Design

The navigation model used in the previous section is based on the classical deliberative sense-plan-act (SPA) control, rooted in a sequential decomposition of the cognitive process in three basic steps; (1) sensing environment and matching the sensed state to a stored world model, (2) planning according to the state, and (3) acting. In fact, most navigation models are based on a centralised model of the environment; a map. Data provided by the robot sensors are used to match the current state to the model in order to plan the path to the goal and to perform related tasks.

Biologically inspired navigation methods that imitate navigational cues observed in animals provide alternatives to map-based navigation strategies (Gelenbe, 1997). In this case, the data obtained from sensors can be used to implement diverse control strategies without either an explicit

Figure 6. Dynamic adaptation of the scanning time T based on the reaction time, Tr, to select individual characters and blocks of information



internal status model or an explicit environment model. Behaviour-based systems are one of the biologically inspired approaches to the navigation problem. The following section shows how a behaviour-based user interface for smart wheelchairs can be designed.

Behaviour-Based Interaction

Behaviour-based systems are a bottom-up reactive approach for the control of real world entities. These systems were originally inspired by biological systems where even the simplest organism interacts effectively with its environment. Behaviour-based navigation consists of determining and maintaining a trajectory to the goal. The main question to be answered in order to navigate is not "Where am I?" but "How do I reach the goal?", and the answer does not always require the initial position to be known. Therefore, the main abilities the robot needs in order to navigate are to move around and to identify goals. Neither a centralised world model nor its position with respect to this model needs to be maintained. The control architecture of these systems is comprised of several lightweight modules, called behaviours. Each behaviour is responsible for reading its own inputs from other behaviours and sensors, and deciding the appropriate outputs to other behaviours and motor actions. The overall system's behaviour emerges from the interaction of those behaviours with the environment.

In shared control navigation systems, such as smart wheelchairs, the high-level behaviours, which are responsible for path-finding methods (Abascal, Lazkano, & Sierra, 2005) are activated by the recognition of the user's purpose and environment status. In order to better recognise the user's intention, the system and the user may share a similar "mental model."

As seen before in traditional navigation systems, the mapping, path planning and driving tasks are achieved using structured environment maps, and the user must maintain an equivalent structured mental map model of the environment in order to select his or her destination. Nevertheless, the user seldom has a structured mental map of the environment due to navigation in unknown buildings or due to his or her physical, and sometimes cognitive, restrictions. This situation suggests that a behaviour-based approach could be more appropriate for these users

When the navigation system and the user interface are based on the behaviour-based approach, the user and the wheelchair share relative navigational concepts, such as "follow the corridor," "find the fire extinguisher," "go to room number 5," "turn left," and so forth. These concepts are easier to process by the user than the traditional ones, making the interaction more natural, similar to the one between humans. Only for feedback purposes, such as informing the user about his or her current position, it may be useful to display a graphical two-dimensional map.

The Interface

Nevertheless, the design of the user interface becomes more complex, because the set of possible behaviours generated by the user can be very large, and it is not easy to detect the current one. For this reason, this kind of system may be appropriate for users that can speak intelligibly. Their orders in natural language can be processed and converted into behaviours. If the users cannot speak, they have to spell the orders by means of an alternative communication system, such as a "one key-scanning system," or choose them from the choices offered in a menu. This procedure may be slow for some navigational tasks. The use of some kind of machine learning mechanism is very useful in dealing with this problem. Learning may improve the entire system performance in a variety of ways: concept generalisation from multiple examples, past experience reuse, new concept discovering for environmental natural landmark learning and, of course, behaviour implementation and coordination. Therefore, learning can be used for improving the interaction of the smart wheelchair both with the environment and with the user.

For instance, recognition of the user's intention can be modelled as a classification problem. Several classes, such as "go out," "dock to the table," "cross the door," "follow the corridor," and so forth, can be set, and then the interaction computed and the user's intention assigned to one of them. This feature also makes it possible to learn from the user during wheelchair navigation, without the supervision of the user, and allows the system to be adapted to the current and changing user needs.

Discussion

The human interface plays the role of translator between the user and the wheelchair controller system. The translation is easier when both intelligent agents, the human and the wheelchair, share a similar vision of the navigational task. Evidently, the design of the interface is largely conditioned by both the user's mental model of the task and the environment, and the wheelchair's navigation model.

The design of the cognitive interaction model is based on a type of concept, well understood on both sides of the interface. This feature allows both complex commands coming from users with a clear mental map of the location and simple commands issued by users with a partial knowledge of the environment, which is not possible for wheelchairs based on classical navigation models.

Whether a classic navigation model is or is not better than a behaviour-based one for the design of smart wheelchairs navigation has not yet been demonstrated. Extensive evaluations and comparisons are still necessary. Nevertheless, as has been discussed, the user-wheelchair interface design is highly conditioned by the underlying navigational model, because an intelligent adaptive user interface has to take advantage of the information handled by the system in order to optimise the dialogue with the user.

CONCLUSION

The design of adaptable mobile interfaces for smart wheelchairs depends heavily on the physical and cognitive characteristics of the user, the tasks that he or she is performing (mainly related to indoor navigation), and the context in which these tasks are being performed. Most of these parameters change quickly, forcing the adaptation of the interface. An intelligent adaptable interface requires modelling the user, the task, and the context in order to be able to compare the current status of the observable parameters and make assumptions that allow it to enhance the communication with the user.

The interface for a smart wheelchair is highly challenging because it requires the optimisation of the communication in order to minimise the quantity of information produced from the user, and hence, the effort required. All of this must be done in such a way that the user maintains control of the main decisions, delegating low-level decisions to the navigation system.

In this chapter, the answers to some of these challenges have been illustrated with examples of the TetraNauta human interface design. The current navigational paradigm of the TetraNauta wheelchair is a classic one, but we are now working on an interface for a behaviour-based navigation model.

ACKNOWLEDGMENT

This work was supported by the Spanish Ministry of Education and Science (MEC) within the projects Heterorred II and III (TIN2004-07994-C04-01 and TIN2005-08976/) and Ambiennet (TIN2006-15617-C03-01/), and by the local government *Gipuzkoako Foru Aldundia*. Borja Bonail's work is supported by a predoctoral fellowship from the Basque Government.

REFERENCES

Abascal, J., Cagigas, D., Garay, N., & Gardeazabal L. (1999). Interfacing users with very severe mobility restrictions with a semi-automatically guided wheelchair. *ACM SIGCAPH Newsletter*, *63*, 16-20.

Abascal, J., Cagigas, D., Garay, N., & Gardeazabal, L. (2002). Mobile interface for a smart wheelchair. In *Human computer interaction with mobile devices* (pp. 373-377), LNCS 2411. Springer.

Abascal, J., Lazkano, E., & Sierra, B. (2005). Behavior-based indoor navigation. In Ambient intelligence for scientific discovery (pp. 277- 300), LNAI 3345. Springer.

Alm, N., Arnott, J. L., & Newell, A. F. (1992). Prediction and conversational momentum in an augmentative communication system. *Communications of the ACM, 35*(5), 46-57.

Baker, M., & Yanco, H. A. (2004, October). Autonomy mode suggestions for improving human-robot interaction. *IEEE Conference on Systems, Man and Cybernetics.*

Butler, C. (1986). Effects of powered mobility on self-initiated behaviors of very young children with locomotor disability. *Developmental Medicine & Child Neurology, 28*, 325-332.

Cagigas, D. (2005). Hierarchical D* algorithm with materialization of costs for robot path planning. *Robotics and Autonomous Systems, 52*, 190-208.

Cagigas D., & Abascal J. (2004). Hierarchical path search with partial materialization of costs for a smart wheelchair. *Journal of Intelligent and Robotic Systems, 39*(4), 409- 431.

Cook, A. M., & Hussey S. M. (2002). AssistivetTechnologies (2nd ed.), In *Principles and practice*. Elsevier.

Ding, D., & Cooper, R .A. (2005). A review of the current state of electric powered wheelchairs. *IEEE Control Systems Magazine*, *25*(2), 22-34.

Drury, J. L., Hestand, D., Yanco, H. A., & Scholtz, J. (2004). Design guidelines for improved human-robot interaction. In *ACM CHI 2004.*

Fehr, L., Langbein, W. E, & Skaar, S. B. (2000). Adequacy of power wheelchair control interfaces for persons with severe disabilities: A clinical survey. *Journal of Rehabilitation Research and Development, 37*(3).

Fernández, J. A., & González, J. (2001). *Multi-hierarchical representation of large-scale space*. The Netherlands: Kluwer Academic Publishers.

Garay, N. (2001). *Sistemas de predicción lingüística. Aplicación a idiomas con alto y bajo grado de flexión, en el ámbito de la Comunicación Alternativa y Aumentativa*. Doctoral dissertation, Servicio Editorial de la Universidad del País Vasco-Euskal Herriko Unibertsitatea.

Garay, N., & Abascal J. (2006). Text prediction systems: A survey. International *Journal on Universal Access in the Information Society (UAIS), 4*(3), 188-203.

Gardeazabal, L. (2000). *Aplicaciones de la Tecnología de Computadores a la mejora de velocidad de comunicación en Sistemas de Comunicación Aumentativa y Alternativa*. Doctoral dissertation, Servicio Editorial de la Universidad del País Vasco-Euskal Herriko Unibertsitatea.

Gelenbe, E. (Ed.) (1997). Special issue: Biologically inspired autonomous systems. *Robotics and Autonomous Systems, 22*(1).

Henlich, O. (1997). Where am I going and how do I get there (an overview of local/personal robot navigation). *Surveys and Presentations in Information Systems Engineering (SURPRISE).* Retrieved August 28, 2006, from http://www.doc.ic.ac.uk/~nd/surprise_97/journal/vol1/oh/

Katevas, N. I., Sgouros, N. M., Tzafestas, S. G., Papakonstantinou, G., Beattie, P., Bishop, J. M., Tsanakas, P., & Koutsouris, D. (1997). The autonomous mobile robot SENARIO: A sensor-aided intelligent navigation system for powered wheelchairs. *IEEE Robotics Automation Magazine*, *4*(4), 60-70.

Latombe, J.C. (1990). *Robot motion planning.* Norwell, MA: Kluwer Academic Publishers.

Levine, S. P., Bell, D. A., Jaros, L. A., Simpson, R. C., Koren, Y., & Borenstein, J. (1999). The NavChair assistive wheelchair navigation system. *IEEE Transactions Rehabilitation Engineering, 7*(4), 443-51.

Myers, B. A., Wobbrock, J. O., Yang, S., Yeung, B., Nichols, J., & Miller, R. (2002). Using handhelds to help people with motor impairments. In *ACM Assets* (pp. 89-96).

Simpson, R. C. (2005). Smart wheelchairs: A literature review. *Journal of Rehabilitation Research & Development*, *42*(4), 423-438.

Simpson, R. C., LoPresti, E. F, Hayashi, S., Guo, S., Ding, D., & Cooper, R. A. (2003). *Smart power assistance module for manual wheelchairs.* 26th International Annual Conference on Assistive Technology for People with Disabilities (RESNA), Atlanta, GA.

Simpson, R. C., Poirot, D., & Baxter, M. F. (2002). The Hephaestus smart wheelchair system. *IEEE Transactions of Neural Systems and Rehabilitation Engineering*, *10*(2), 118-22.

Vicente-Díaz, S., Amaya-Rodríguez, C., Díaz-del-Río, F., Civit-Balcells, A., & Cagigas, D. (2002). TetraNauta: An intelligent wheelchair for users with very severe mobility restrictions. In *Proceedings of the 2002 IEEE International Conference on Control Applications* (pp. 778-783).

Yanco, H. A. (1998). Wheelesley, a robotic wheelchair system: Indoor navigation and user interface. In V. O. Mittal, H. A. Yanco, J. M. Aronis, & R. C. Simpson (Eds.), *Assistive technology and artificial intelligence* (pp. 256-268). Springer-Verlag.

Yanco, H. A., & Drury, J. L. (2004, October). Classifying human-robot interaction: An updated taxonomy. In *Proceedings of the IEEE Conference on Systems, Man and Cybernetics.*

Yanco, H. A., & Gips, J. (1998). Driver performance using single switch scanning with a powered wheelchair: Robotic assisted control versus traditional control. In *Proceedings of the Annual Conference of the Rehabilitation Engineering and Assistive Technology Society of North America*, RESNA Press (pp. 298-300).

KEY TERMS

Adaptive Interfaces: Intelligent user interfaces able to adjust their operation to the current situation, based on the actual values of diverse parameters which are defined in models.

Input Device: Any machine that feeds data into a computer. In this context, input devices are typically the appliances handled by humans to interact with computers.

Modelling: A methodology that allows representing the available knowledge about users, tasks, context, and so forth, by means of relevant and observable parameters.

Navigation: The set of procedures followed by mobile vehicles or people to select a trajectory from one point to other and to move following it.

Shared Control: A robotics control paradigm where the navigation tasks are developed by both the system and the user, in a collaborative fashion.

Smart Wheelchairs: Autonomous mobile vehicles devoted to improve the mobility of users with severe motor impairments that experience difficulties in driving traditional electric wheelchairs. A smart wheelchair typically consists of a standard powered wheelchair to which a collection of sensors and a computer (for control purposes) have been added.

Word Prediction: Consist of a set of techniques based on diverse methods, such as statistics or natural language processing, that allow the system to anticipate the word that the user is trying to input.

ENDNOTES

1. AAC systems substitute or complement the oral communication. The International Society for Augmentative and Alternative Communication (ISAAC) is the best source of information on Alternative communication: http://www.isaac-online.org
2. The users were people affected by paraplegia and tetraplegia admitted to the Hospital Nacional de Parapléjicos de Toledo (Spain).
3. Multi-sensor data fusion seeks to combine information from multiple sources to reach inferences that are not feasible from a single sensor. This task is not trivial, as sensor outputs often have overlaps and conflicts, their location is usually highly distributed, and their configuration very dynamic. In addition, their performance can vary with time. There are diverse techniques for the implementation of data fusion models, based on classical statistics with a mathematical background which guarantees soundness.
4. The extensive study on Assistive Technologies by Cook and Hussey, 2002, can be consulted for more information about input systems.
5. This kind of input system arranges the set of options in one or more matrices that are row-column scanned. When the desired item is emphasised, the user selects it by means of the activation system, usually a switch handled by means of any residual controlled movement (in the hand, head, breathing, sound, etc.).
6. Alm et al., 1992, estimated that 10 words per minute is the highest communication rate which can be obtained with this kind of system.

Section IV
Evaluation Techniques for Mobile Technologies

The rapid evolution of mobile technologies has posed a number of challenges in terms of effective evaluation strategies, which has opened this area up to interesting, ongoing debate, especially with regards the value of lab versus field evaluations. Starting with a theoretical look at the concepts and issues involved in evaluating mobile human-computer interaction, this section covers the spectrum of evaluation as it applies to mobile technologies. Chapters are included which look at adaptation of traditional methods to meet the needs of mobile evaluations, and chapters that outline means to systematically select, combine, and tailor methods to the specific needs of any given evaluation. The use of appropriated heuristic evaluation, wizard-of-oz studies, cognitive modeling as a testing tool, and Fitt's Law as a performance measure are all covered in this section. The use of multilayered evaluation approaches, the application of the Privacy Regulation Model as an evaluation tool, and a framework and model for identifying, organising, and classifying usability factors of mobile phones are also discussed. This section reflects the aforementioned debate regarding lab vs. field evaluation of mobile technologies by including chapters representing both sides of the argument, including those which present innovative mechanisms and set-ups for use in each context.

Chapter XLIII
Evaluating Mobile Human-Computer Interaction

Chris Baber
The University of Birmingham, UK

ABSTRACT

In this chapter the evaluation of human computer interaction (HCI) with mobile technologies is considered. The ISO 9241 notion of 'context of use' helps to define evaluation in terms of the 'fitness-for-purpose' of a given device to perform given tasks by given users in given environments. It is suggested that conventional notions of usability can be useful for considering some aspects of the design of displays and interaction devices, but that additional approaches are needed to fully understand the use of mobile technologies. These additional approaches involve dual-task studies in which the device is used whilst performing some other activity, and subjective evaluation on the impact of the technology on the person.

INTRODUCTION

This chapter assumes that 'usability' is not a feature of a product, that is, it does not make sense to call a product itself 'usable'. Rather, usability is the consequence of a given user employing a given product to perform a given activity in a given environment. Holcomb and Tharp (1991) proposed a 'model' of interface usability, which is illustrated by Table 1. The definitions presented in Table 1 arose from consideration of the user interface of desk-based computers. However, it ought to be apparent that the majority of the components are defined in terms of an individual's perceptions of features of the user interface.

The International Standards Organization has a number of standards relevant to human-computer interaction (Bevan, 2001). Current standards for mobile devices tend to focus on product attributes, for example, *ISO 18021: Information Technology —User Interface for Mobiles* (2001) provides interface specifications for Personal Digital Assistants. Other Standards have recognized the multifaceted nature of usability and have sought to encourage an approach that is similar to Quality Assessment (Earthey et al., 2001). Demonstrating compliance

Table 1. Holcomb and Tharp's (1991) "model" of interface usability

Component	Term
Functional	Able to accomplish tasks for which software is intended Perform tasks reliably and without errors
Consistent	Consistent key definitions Show similar information at same place on screens Uniform command syntax
Natural and Intuitive	Learnable through natural conceptual model Familiar terms and natural language
Minimal memorization	Provide status information Don't require information entered once to be re-entered Provide lists of choices and allow picking from the lists Provide default values for input fields
Feedback	Prompt before destructive operations like DELETE Show icons and other visual indicators Immediate problem and error notification Messages that provide specific instructions for action
User help	Online help system available Informative, written documentation
User control	Ability to undo results of prior commands Ability to re-order or cancel tasks Allow operating system actions to be performed within the interface

with the standards requires analysts to document their evaluation, demonstrating how it meets the objectives of the standard. The definition of usability offered by the International Standards Organization, that is, in ISO9241, part 11, is, "… *the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use.*" (ISO9241-11, 1998). The implications are that, first, usability is the consequence of a given user employing a given product to perform a given activity in a given environment (as stated) and, second, that it is possible to measure aspects of this relationship in terms of effectiveness, efficiency, and user satisfaction. It is important to note that these three aspects are inter-connected and that any evaluation activity ought to try to measure some aspect of each (Frøkjær et al., 2000).

Defining Evaluation Targets

If one is able to speak of measures, then it makes sense to be able to determine some criteria that indicate good or poor performance on these measures. Good et al. (1986) proposed that it is important to define both evaluation targets and metrics that relate to these targets. For example, in a study of conferencing systems, Whiteside et al. (1988) identified 10 attributes that they felt reflected the use of the conferencing system, for example, ranging from a fear of feeling foolish to a number of errors made during task performance. For each attribute, Whiteside et al. (1988) defined a method for collecting data about that attribute, for example, questionnaires, observation, and so forth, and then set performance limits relating to best, worst, and planned levels. A study of a wearable computer for paramedics (Baber et al., 1999) used this concept to produce Table 2. In Table 2, three measures of performance were undertaken, that is, predictive modeling (using critical path analysis), user trials, and performance improvement arising from practice. In addition, three subjective evaluation methods were used. Table 2 shows how the system met (or exceeded) some of the target criteria but fell below the target for time (although it is not within the 'worst' case range). One benefit of such a technique is to allow

Table 2. Defining evaluation targets

Factors	Method	Metrics	Worst	Target	Best	Current
Performance						
Task	CPA	Time	-15%	0	+5%	-2%
Practice	User trials 1st vs. 3rd trial	Time % change	-15% 1st > 3rd	0 3rd > 1st	+5% 0	-10% 3rd>1st
Subjective evaluation	SUS[1] SUMI[2] Heuristics	Scale: 0-100 Scale: 0-100 Scale: 0-10	50 50 <6	60 60 6	70 70 >6	65 60 8

the design team to decide whether there is a need for more effort to refine the device, or whether, having met (some or all of) the requirements, the design process can be closed.

The ISO9241 notion of usability requires the concept of evaluation targets, for example, one could begin with a target of "66% of the specified users would be able to use the 10 main functions of product X after a 30 minute introduction." Once this target has been met, the design team might want to increase one of the variables, for example, 85% of the specified users, or 20 main functions, or 15 minute introduction, or might want to sign-off that target.

Why Conduct Evaluation?

The concept of usability that is used in this chapter (and in ISO9241) implies that changing any one of the variables {user, activity, device, environment} can have an impact on usability. This implication points to the well-known assertion that an activity that a designer of the product might find easy to perform could prove problematic for a user who has had little or no previous experience of the product. It also points to potential issues relating to the usability of mobile technology, particularly through consideration of the environment. If we think about sending a text-message from a handheld device, such a mobile telephone or a Blackberry™, the activities involved could be somewhat different while sitting on a train versus walking down a busy street. This change in environmental setting will have a marked effect on usability of the device. This does not necessarily result from the design of the device itself but rather from the interactions between design, use, and environment. As Johnson (1998) pointed out, *"HCI methods, models and techniques will need to be reconsidered if they are to address the concerns of interaction on the move."* (Johnson, 1998). The question for this chapter, therefore, is how best to address the relationship between user, activity, product, and environment in order to evaluate the usability of mobile technology. Related to this question is how evaluation might capture and measure this relationship, and then what can designers do to improve usability. This latter point is particularly problematic if one assumes that design is about creating a product rather than about creating an interaction.

Before considering these questions, it is worth rehearsing why one might wish to conduct evaluation. Baber (2005) notes that the primary reason for conducting evaluation, in HCI, is to influence design (ideally, to *improve* the product). This implies that evaluation ought never to be a one-off activity to be conducted at the end of the design lifecycle in order to allow a design to be signed-off (Gould & Lewis, 1985; Johnson, 1992). Rather, it means the following:

1. Evaluation is a recursive activity that cuts across the entire design lifecycle, for example, software engineers will run versions of the code to debug and check; product designers will continually critique and refine their concepts. What is not always apparent is the

manner in which these processes could (or indeed ought) be made formal and to result in something that can be communicated to other members of the design team.

2. Evaluation should be incorporated into as many stages of design as possible—this points to (i) but also raises that questions of recording and communicating the results of evaluation in a many that can be beneficial to the design process.
3. Evaluation should be designed to maximize the impact of the evaluation of the design stage in which it is used—the suggestion is that, rather engaging in evaluation as a mandated exercise to allow sign-off between stages, it ought to be an activity that positively advances the design process.
4. Evaluation should guide and inform design activity—the results of any evaluation should be reported in a manner that can lead to change in the design and can be reported in a manner that is transparent and reliable.

A final point to note is that evaluation is a process of comparing the product against something else, for example, other products, design targets, requirements, standards. Thus, evaluation requires a referent model (Baber, 2005). It is naïve to believe that one can "evaluate" something in a vacuum, that is, to think that one can take a single product and "evaluate" it only in terms of itself. In many ways this is akin the concept of a control condition in experimental design; one might be able to measure performance, but without knowing what would constitute a baseline for the measure, it is not possible to determine whether it is good or bad.

Defining Referent Models

While it might be fairly clear as to why comparison requires a referent model, there is a problem for novel technologies. After all, the point of these technologies is to move beyond the conventional desk-bound personal computers and this will ultimately create new forms of interaction. However, the move to different technologies makes it hard to establish a sensible basis for evaluation. What is the referent model for mobile HCI?

A common form of mobile technology is the digital tour-guide that, knowing where the user is (using Global Positioning Satellite (GPS) to determine location) and what the user is doing, can provide up-to-the-minute information to help the user. There are few, if any, products that are like these concepts, so what constitutes a referent? At one level, this is simply because future HCI is attempting to develop approaches to interaction with technology for which there are no existing models. The answer to this question, the author suggests, comes from the assertion at the start of this chapter: usability is not a characteristic of the product, but the result of the interactions between user, product, activity, and environment. If we assume that tourists have a variety of strategies and artifacts that they currently use to find out where they are or to find out interesting information about a particular location, for example, maps, books, leaflets, other people. One could ground an initial evaluation of using the product to perform a given set of activities in comparison with existing practices. Conducting evaluation against other products in terms of a set of activities offers the analyst the following benefits:

1. The evaluation will cover a range of functions on the products. It is important to ensure that the comparison provides a fair and accurate view of the product. After all, it is not really the point of evaluation to just demonstrate the product X is better than product Y—partly because there are bound to be occasions when products X and Y are similar, or where product Y is better than product X, and partly because simply knowing that X > Y tells us very little about how to improve X (or Y) or why X is superior.
2. The focus of the evaluation is less on product functioning than on user activity. This might appear, at first glance, to be tautological—surely product evaluation is about evaluating the product? This is, of course, true in a technical sense. However, HCI is about human-computer *interaction*, and

the defining feature of this relationship is the interaction (rather than either human or computer). If one is concerned with technical evaluation then, perhaps some of the features to be included in a comparison table (like the one shown in Table 2) would be some of the technical features, for example, processor speed, RAM, memory, and so forth.

3. As the evaluation is concerned with user activity (as opposed to product functioning), the type of metrics that could be applied may well change. When comparing user activity on two or more products, it is important to decide what information is really being sought. Do we want to know only that X>Y? Or do we want to know that using product X or Y have differing effects on user activity?

In the field of mobile and wearable computers, much of the evaluation research has focused on comparing performance on a wearable computer with performance using other media. Thus, studies might compare performance using a wearable computer, say to perform a task that involves following instructions, and find that sometimes performance is superior in the paper condition (Siegel & Bauer, 1997; Baber et al., 1999) and sometimes it is superior in the wearable computer condition (Bass et al., 1995, 1997; Baber et al., 1998). This highlights the potential problem of comparing disparate technologies in an evaluation; it is not clear that any differences in performance are due to the experiment favoring one technology over another or whether there are other factors at play here. For example, a common observation is that people using the wearable computer tend to follow the instructions laid out on the display, whereas people using paper tend to adopt a more flexible approach (Siegel & Bauer, 1997; Baber et al., 1999). The notion that technology influences the ways in which people work is often taken as 'commonsense' by Human Factors engineers. However, the question of how and why such changes arise ought to have a far deeper impact on evaluation than is currently the case. As mentioned earlier, one way to deal with this problem is to focus on activities that people are performing using a variety of products. However, this will only cope with part of the problem. For instance, the electronic tour-guide given could be evaluated in comparison with other ways of performing activities, but this does not tell us whether any differences between the electronic tour-guide and the other products are due to the concept or to the realization of the concept or to changes in the activity arising from the use of the device. In other words, if we find that the electronic tour-guide performs less well than speaking to someone, is this because the tour-guide lacks information, or because it lacks clear presentation of information, or because it lacks speedy access to the information, or because it lacks flexibility of response, or because of some other reason (the evaluation could point to all of these, not to specific reasons).

At one level, the evaluation of mobile HCI calls for the application of current evaluation techniques. However, there are other aspects of future HCI that call for rethinking of evaluation. In other words, it might not be entirely appropriate to take methods that have proven useful for evaluating desktop HCI and apply these to future HCI. As Wilson and Nicholls (2002) point out in discussing the evaluation of virtual environments:

There are only a limited number of ways in which we can assess people's performance...We can measure the outcome of what they have done, we can observe them doing it, we can measure the effects on them of doing it or we can ask them about either the behavior or its consequences. (Wilson & Nicholls, 2002)

The underlying assumption here is that human behavior is measurable in a finite number of ways. Combining this assertion with the need to study the relationship between user, activity, device, and environment, it becomes apparent that evaluation of user interaction with mobile activity can be reduced to a small number of requirements. Furthermore, the ISO 9241 notions of efficiency, effectiveness, and satisfaction point to the approaches outlined by Wilson and Nicholls (2002). For example, efficiency could be considered in terms of the amount of resource expended in order to achieve a goal (perhaps in terms of time to complete a task), and effectiveness could relate to the quality of this performance (perhaps in terms of the amount of activity completed or the quality of the outcome), and satisfaction would relate to the user perception of the activity (perhaps in terms of a judgment relating to their own performance or effort, perhaps relating to some aspect of using the device). What is required is not so much a battery of new measures, so much as an adaptation of existing approaches that pay particular attention to the relatively novel aspects of the environment and activity that pertain to mobile devices.

MAKING SENSE OF HUMAN ACTIVITY WITH MOBILE TECHNOLOGY

The argument so far is that what needs to be evaluated is not simply the product, but the interaction between user, activity, device, and environment. This raises the question of what can be defined as appropriate forms of activity. The first issue for mobile technology is the assumption that it is to be used on the move, which raises two possibilities: (1) 'on the move' means physically moving, for example, walking, driving a car, traveling as a passenger; (2) 'on the move' means being in different places away from 'normal' office environments. One problem relating to both of these possibilities is the difficulty of collecting data in the field—there are problems arising from recording the data, managing the collection of data, and controlling experimental conditions that are far from trivial. However, if evaluation studies involve managing the interactions between user, activity, device, and environment, then it might not be possible to concentrate efforts on specific aspects of the interactions, for example, comparing the use of the device under different mobility conditions.

Interacting with Mobile Technology while Walking

Consideration of interaction while moving immediately suggests that asking people to evaluate a product whilst sitting down in a laboratory might lead to different results than when using the product 'on the move'. This is just what Kjeldskov and Stage (2004) demonstrated. Indeed, they found that having participants report usability problems while sitting down in the laboratory led to more usability problems being reported than when the participants performed the evaluation while walking. They suggested that this result might have arisen from different demands on attention—in the seated condition there was little distraction from the product and so participants were able to devote most of their attention to it, but in the walking conditions, attention needed to be divided between the device and the task of walking. This effect can be compounded by variation in other contextual factors, such as lighting levels and complexity of the path that one is following (Barnard et al., 2007).

It has been demonstrated that walking can impact cognitive tasks (Ebersbach et al., 1995), and so the use of a mobile device could be thought of in terms of a 'dual-task'. A common methodological approach in Ergonomics/Human Factors involves asking participants to perform one task while attending to another, for example, tracking a line on a screen while also performing mental arithmetic. There are several reasons why this approach is useful, both in terms of developing theory of human performance and in terms of considering how combinations of tasks can be modified. In broad terms, the assumption is that the human ability to process information from several sources can be compromised under conditions of increasing complexity. Complexity might arise from the difficulty of one or both of the tasks, from the quality of the signals being attended to, from the amount of interference between the two tasks, and so forth. By measuring performance on the tasks under different levels of complexity, it is possible to judge the person's ability to perform and the amount of interference that could occur.

Taking the dual-task paradigm as a starting point, one can consider many forms of mobile technology to be used not only in different places but also while the person is physically moving, for example, walking down a busy street or following a predefined route or walking on a treadmill. Thus, one approach to studying mobile technology from a dual-task perspective would involve measuring some aspect of walking and some aspect of using the technology. Barnard et al. (2005) compared reading tasks on a Personal Digital Assistant (PDA) while walking on a treadmill and walking along a defined path. They found a reduction in walking speed (by around 33%) compared to walking without performing the tasks on the device. This indicates that using the device leads to measurable changes in walking activity. They found no difference in comprehension between conditions (although it is often difficult to find measurable differences in comprehension in experiments that involve reading from screens, see Dillon, 1992), but they did find that word search took significantly longer when walking along the path than on the treadmill. This result suggests that participants walking the path had more need to divide their attention between the world and the device, and indeed, path following correlated with the use of scroll bars on the device, suggesting that more attention on the world led to more need to scroll the text to find one's place while reading. What is interesting about this particular study is that it reports objective results on both primary (using the device) and secondary (walking under different conditions) tasks, and shows some interactions between the two. This shows that modifying the environment has a bearing on activity which, in turn, affects user performance (even with the same device requiring the same activity).

Using Mobile Technology while On-the-Move

Prototypical mobile technologies often address scenarios related to tourists because this emphasizes the need to move around an unfamiliar environment and the desire for information relating to the world around us, for example, routes to places, interesting information about landmarks, advice on traveling, and so forth. Considering the scenario from the perspective of usability, evaluation could allow the design team to agree on 'benchmark' levels of performance using existing practices, and to then consider what benefits might accrue from modifying those practices through the introduction of technology.

While the activity of walking can interact with the use of the mobile device, there are other aspects of use on-the-move that can also play a role. Duh et al. (2006) compared the evaluation of a mobile telephone, used to perform a set of activities, in the laboratory and on a Mass Rapid Transit (MRT) train in Singapore. The study showed that participants encountered significantly more problems in the train condition than in the laboratory. The authors relate the problem to five primary areas: ambient noise levels, movement of the train, issues relating to privacy, increase in effort needed to perform the activity, additional stress, and nervousness. Of these factors, the main ones relate to aspects of the environment, viz. noise and movement of the train, and these, in turn, have a bearing of the ability of participants to complete the activity. In addition to the affect of movement on the performance of the users, the impact on the performance of the technology is equally important, for example, what happens when wireless networks do not cover the whole of the area and the user encounters 'shadows', or what happens when positioning systems have drift or inaccuracies. One approach might be to attempt to guarantee optimal delivery of service at all times by modifying the infrastructure rather than the device, for example, with boosters located in the environment. Another approach would be to provide ways of informing the user about accuracy of the data on which the system is working (Bell et al., 2006).

Devices can also be used while driving automobiles and there is evidence that interference between the activity of driving and the activity of using the device are more than simply physical (Boase et al., 1988; Brookhuis et al., 1991; Svenson & Patten, 2005; Wikman et al., 1998). This means that using a 'hands-free' kit will not eliminate all forms of interference. For example, Nunes and Recarte (2002) show that the more cognitively demanding a telephone conversation, the greater the reduction in the user's ability to attend to the environment while driving. This research further highlights the problem of isolating the usability of the device itself from the interactions between user, activity, and environment.

SUBJECTIVE EVALUATION OF TECHNOLOGY

"[U]ltimately it is the users of a software system [or any product] who decide how easy its user interface is to manipulate…" (Holcomb & Tharp, 1991). Thus, one might feel that asking people about the product would be the obvious and most useful approach to take. However, there are several problems with this approach, for example, people might not always be able to articulate how they feel about the product (so the reports might be incomplete or inconsistent), people might use a variety of previous experiences as their referent models (so it might be difficult to generalize results across respondents), people might not be able to respond critically to the product (so there might be a 'halo-effect' with the participant responding to the novelty of the device rather than considering issues of usability). For these and other reasons, it is common practice to provide some structure to subjective evaluation, usually through some form of procedure or checklist. Furthermore, it would be suggested that subjective evaluation should be used as a secondary measure as far as practicable, with the primary focus on data collected from user trials.

Subjective Response to the Device

Participants could be asked to walk-through the performance of a given activity using the device, by explaining what they are doing and why. Monk et al. (1986) presented a detailed set of guidelines on how to use walk-through approaches to the evaluation in their Cooperative Evaluation method. The main aim of the approach is to capture problems that users experience when using a product.

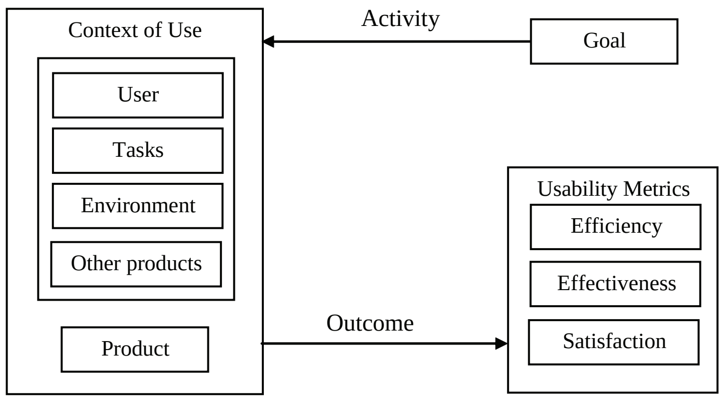
In terms of checklists, a great deal of research effort from the late 1980s to the mid 1990s led to the development of a number of usability surveys. Some, like CUSI-Computer User Satisfaction Inventory (Kirakowski & Corbett, 1988) and QUIS-Questionnaire for User Interface Satisfaction (Chin et al., 1988), are designed to capture user response to an interface, particularly in terms of affective components (such as satisfaction). Others, like the checklist of Ravden and Johnson (1989) or SUS (Brooke, 1996), have been designed to cover both aspects of the interface and characteristics of usability. While these surveys are based on sound HCI principles, interpretation is left to the analyst who could lead to potential bias or misinterpretation. The SUMI checklist (Kirakowski, 1996) was developed using a rigorous approach to defining appropriate components of usability and presents results in terms of a comparison with a database of previous evaluations.

Subjective Responses to Using the Device To Perform an Activity

In addition to eliciting opinions from users regarding the device, researchers are also keen to obtain reactions of some of the consequences of using the device. By way of analogy, if we consider the virtual reality research community, we can see efforts to elicit reaction to either the physical effects of using virtual reality, for example, Cobb et al.'s (1999) Virtual Reality Induced Symptoms and Effects (VRISE) or the measurement of 'presence'(Slater et al., 1994; Witmer & Singer, 1998). In the domain of wearable computers, physical effects have been evaluated using self-report on a comfort rating scale (Knight et al., 2002).

In terms of performing an activity, researchers often make use of the NASA-TLX (Hart & Staveland, 1988) which measure subjective response to workload. The basic notion is that activities make different demands on people in terms of time pressure or mental effort, and can lead to different responses such as frustration or perceived level of performance. The NASA-TLX captures these responses and can be used to compare perceptions of users with combinations of different devices or activities.

Figure 1. ISO9241 usability evaluation process



DESIGNING AN EVALUATION PROTOCOL

Throughout this chapter, emphasis has been placed on the notion of 'context of use' and the concept of usability defined by ISO 9241, pt. 11. The relationship between these concepts is illustrated by Figure 1. In order to evaluate any item of technology, one needs to plan an appropriate campaign of evaluation—this means consideration of the evaluation from the start of the design process and performance of evaluation as often as practicable during the course of development. Assume that, whatever design process is being followed, there will be four primary phases: initial concept development, prototyping, specification and build. At each phase, the form and type of evaluation will change (depending on access to functionality on the product as much as anything else), but the basic considerations remain constant, that is, adequately defining context of use and applying appropriate usability metrics.

Before elaborating on Figure 1 as a process, it is worth re-emphasizing the point made earlier that usability evaluation always involves comparison with the product being considered against some referent model. The referent model could be other products, but is equally likely to be a set of design targets (see Table 2). In terms of comparison, a set of usability metrics can be applied. The 'efficiency' metric relates to the manner in which resources are applied during the activity in order to achieve the outcome; the 'effectiveness' metric relates to the completion of the outcome; the 'satisfaction' metric relates to the user's response to performing the activity. Needless to say, all metrics apply to a given user performing a given activity in order to achieve a given goal in a given context of use with a given product. In terms of what to measure, each metric has several options. For the sake of brevity, in this chapter, one quantitative and one qualitative measure for each metric will be considered (the reader is encouraged to review ISO 9241, pt. 11 as a starting point for considering alternatives). For 'efficiency', a quantitative metric could be the number of mistakes a person made when using the product, and a qualitative metric could be a subjective workload (using the NASA-TLX mentioned); for 'effectiveness', a quantitative metric could be time to achieve the goal and a qualitative metric could be a subjective rating of performance; for 'satisfaction', a quantitative metric could be time spent using the device (over the course of several days) and a qualitative metric could be a self-report of how pleasant the product was to use. It should be apparent that the distinction between efficiency, effectiveness, and satisfaction is somewhat arbitrary, which is why is important to make sure that all three metrics are applied during evaluation.

The idea that evaluation requires a 'protocol' is meant to imply that one ought to approach it in much the same way that one approaches the design of an experiment, that is, by defining independent variables, which are the goal, activity, and context of use, and by defining dependent variables, which are the usability metrics. The notion of the referent model also makes sense in terms of experimental design because the 'hypothesis' under test is that the outcome will be equal to or better than the referent model.

Initial Concept Development

During the 'initial concept development' phase, it is possible that one of the components of 'context of use' will dominate the others. For example, a designer might have an idea about how the product will function or how to perform a particular task or how to help a particular user. In order to explore this concept, designers make use of scenarios in various forms, for example, storyboarding, sketching, rich pictures, illustrative stories, and so forth. From Figure 1, it can be argued that a good scenario would include (as a minimum) some consideration of the type of person who would be likely to use the product, the tasks that the person would perform in order to achieve specific goals (as well as any other tasks that might need to be performed concurrently), the environment in which they might be performing these tasks, and the presence or use of other products to support this activity. In the domain of 'traditional' computer systems, the 'environment' can be assumed to be more or less constant, that is, the computer would be used on a desk in an office. In mobile computing, the 'environment' will have a significant impact on how the product will be used, as will the range of tasks that the person will be performing. It is this impact of the environment and the increasing range of concurrent tasks that makes evaluating mobile technology different from other computer applications. One way in which these aspects can be considered is to develop a scenario in which a person achieves the defined goal using no technology, another in which they use 'contemporary' technology and another in which they use the concept product. By storyboarding these different scenarios, the design team gets a feeling for the main benefits to be gained from using the product (and an appreciation as to whether or not to pursue its development). During this stage, the usability metrics can be defined in terms of what measures can sensibly differentiate the product from any alternative ways of performing the task.

Prototyping

During 'prototyping' different versions of the product are developed and tested. The prototype need not be a fully-functioning product. Indeed, Nilsson et al. (2000) shows how very simple models can be used to elicit user responses and behaviors. Their study involved the development of a handheld device (the 'pucketizer' for use in water treatment plants and the initial studies had operators walking around the plant with a non-functioning object to simulate the device. From this experience, the design team went on to implement a functioning prototype based on an 8-bit microcontroller, wireless communications, and a host computer running a JAVA application). This work is interesting because it illustrates how embedding the evaluation process in the environment and incorporating representative end-users lead to insights for the design team. Taking this idea further, it is feasible for very early prototyping to be based on paper versions. For example, one might take the form factor of the intended device (say a piece of wood measuring 5" x 3" x ½"—which is approximately the size of Personal Digital Assistant) and then placing 3" x 2" paper 'overlays' to represent different screen states—change the 'screens' is then a matter of the user interacting with buttons on the 'product' and the evaluator making appropriate responses. Of course, this could be done just as easily using an application in WinCE (or through the use of a slideshow on the device), but the point is that initial concepts can be explored well before any code is written or any hardware built.

Specification and Build

'Specification and build' is the phase that one might traditionally associate with evaluation. Evaluation activity at this phase of the design process would be 'summative' (i.e., occur at the summation of the process), and would usually be used to confirm that the design was acceptable prior to committing to manufacture. At this stage, the main concerns regarding hardware and software would have been dealt with and so any usability evaluation that would call for significant change to hardware or software is likely to be ignored (unless the product is scrapped and the process started again, or unless these recommendations are filed for the next version of the product). However, usability evaluation can play an important role in this phase because it will form part of the acceptance testing of end-users and could, if positive, play a role in defining marketing activity or informing training requirements.

CONCLUSION

While the concept of usability as multi-faceted might seem straightforward, it raises difficult problems for the design team. The design team focuses its attention on the device, but the concept of usability used in this chapter implies that the device is only part of the equation and that other factors relating to the user and environment can play significant roles. The problem with this, of course, is that these factors lie outside the remit of the design team. One irony of this is that a well-designed device can 'fail' as the result of unanticipated activity, user characteristics, and environmental features.

The issue raised in this chapter is that evaluating mobile technology involves a clear appreciation of the concept of usability, in line with ISO standard definitions. The ISO9241 concept of usability emphasizes the need to clearly articulate the 'context of use' of the device, through consideration of user, activity, device, and environment. This means that evaluation has to take account of the interactions between user, activity, device, and environment. What is essential is that evaluation is conducted in a way that ensures a good fit between the 'context of use' in the real-world and that simulated in the laboratory. This does not mean that one needs to include all aspects of the real-world in the laboratory but that one is able to reflect key variables and that the evaluation is designed to ensure a balanced comparison. It would be easy to 'prove' that a given device was superior to any other device simply by ensuring that the test favored the device in question. It is equally easy to 'prove' that evaluation in the 'laboratory' do not reflect performance in the 'real-world'. However, such studies often reflect a limited grasp of adequate experimental design and, ultimately, a poor understanding of science. One is not 'proving' that a product is well designed through evaluation. Rather one is demonstrating 'fitness-for-purpose' under a well-defined context of use.

REFERENCES

Baber, C. (2005). Evaluation of human-computer interaction. In J. R. Wilson & E. N. Corlett (Eds.), *Evaluation of human work* (pp. 357-388). London: Taylor and Francis.

Baber, C., Arvanitis, T. N., Haniff, D. J., & Buckley, R. (1999). A wearable computer for paramedics: Studies in model-based, user-centered and industrial design. In M. A. Sasse & C. Johnson (Eds.), *Interact'99* (pp. 126-132). Amsterdam: IOS Press.

Baber, C., Haniff, D. J., Knight, J., Cooper, L., & Mellor, B. A. (1998). Preliminary investigations into the use of wearable computers. In R. Winder (Ed.), *People and computers XIII* (pp. 313-326). Berlin: Springer-Verlag.

Baber, C., Haniff, D. J., & Woolley, S. I. (1999). Contrasting paradigms for the development of wearable computers. *IBM Systems Journal, 38*(4), 551-565.

Barnard, L., Yi, J. S., Jacko, J. A., & Sears, A. (2005). An empirical comparison of use in- motion evaluation scenarios for mobile computing devices. *International Journal of Human-Computer Studies, 62,* 487-520.

Barnard, L., Yi, J. S., Jacko, J. A., & Sears, A. (2007). Capturing the effects of context on human performance in mobile computing systems. *Personal and Ubiquitous Computing, 11*(2), 81-96.

Bass, L., Kasabach, C., Martin, R., Siewiorek, D., Smailagic, A., & Stivoric, J. (1997). The design of a wearable computer. In *Proceedings of the CHI '97* (pp. 139-146). New York: ACM.

Bass, L., Siewiorek, D., Smailagic, A., & Stivoric, J. (1995). On site wearable computer system. In *Proceedings of the CHI '95* (pp. 83-88). New York: ACM.

Bell, M., Chalmers, M., Barkhuus, L., Hall, M., Sherwood, S., Tennent, P., Brown, B., Rowland, D., Benford, S., Hampshire, A., & Capra, M. (2006). Interweaving mobile games with everyday life. In *Proceedings of the CHI 2006* (pp. 417-426). New York: ACM.

Bevan, N. (2001). International standards for HCI and usability. *International Journal of Human Computer Interaction, 55*, 533-552.

Boase, M., Hannigan, S., & Porter, J. M. (1988). Sorry, can't talk now…just overtaking a lorry: The definition and experimental investigation of the problem of driving and hands-free car phone use. In E. D. Megaw (Ed.), *Contemporary ergonomics.* London: Taylor and Francis.

Brooke, J., (1996). SUS: a quick and dirty usability scale. In P. W. Jordan, B. Weerdmeester, B. A. Thomas, & I. L. McLelland (Eds.), *Usability evaluation in industry* (pp. 189-194). London: Taylor and Francis.

Brookhuis et al, (1991). The effects of mobile telephoning on driving performance. *Accident Analysis and Prevention, 23*

Chin, J. P., Diehl, V. A., & Norman, K. L. (1988). Development of an instrument measuring user satisfaction of the human-computer interface In *Proceedings of the CHI '88* (pp. 213-218). New York: ACM.

Cobb, S. V. G., Nichols, S. C., Ramsey, A. D., & Wilson, J. R. (1999). Virtual reality induced symptoms and effects (VRISE). *Presence: Teleoperators and Virtual Environments, 8*, 169-186.

Dillon, A. (1992). *Designing usable electronic text.* London: Taylor and Francis.

Duh, H. B-L., Tan, G. C. B., & Chen, V. H-H. (2006). Usability evaluation for mobile devices: a comparison of laboratory and field tests. In *Proceedings of Mobile HCI '06* (pp. 181-186). Berlin: Springer-Verlag.

Earthey, J., Sherwood Jones, B., & Bevan, N. (2001). The improvement of human-centered processes—Facing the challenges and reaping the benefit of ISO 13407. *International Journal of Human Computer Studies, 55,* 553-585.

Ebersbach, G., Dimitrijrvic, M. R., & Poewe, W. (1995). Influence of concurrent tasks on gait: A dual task approach. *Perceptual and Motor Skills, 81,* 107-113.

Frøkjaer, E., Hertzum, M., & Hornbaek, K. (2000). Measuring usability: are effectiveness, efficiency and satisfaction really correlated? In *Proceedings of the CHI '2000* (pp. 345-352). New York: ACM.

Good, M., Spine, T. M., Whiteside, J., & George, P. (1986). User-derived impact analysis as a tool for usability engineering. In *Proceedings of the CHI '86* (pp. 241-246). New York: ACM.

Gould, J. D., & Lewis, C. H. (1985). Designing for usability: Key principles and what designers think. *Communications of the ACM, 28*(3), 300-311.

Gray, W. D., & Salzman, M. C. (1998). Damaged merchandise? A review of experiments that compare usability evaluation methods. *Human-Computer Interaction, 13*(3), 203-261.

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P. A. Hancock & N. Meshkati (Eds.), *Handbook of mental workload.* Amsterdam: North Holland.

Holcomb, R., & Tharp, A. L. (1991). What users say about software usability. *International Journal of Human-Computer Interaction, 3*(1), 49-78.

ISO 18021. (2001). Information technology—User interface for mobiles. ISO9241-11 1998.

ISO 9241. (1998). Ergonomics of office work with VDTs—Guidance on usability. Geneva: International Standards Office.

Jeffries, R., Miller, J. R., Wharton, C., & Uyeda, K. M. (1991). User interface evaluation in the real world: A comparison of four techniques *(CHI'91)* (pp. 119-124). New York: ACM.

Johnson, P. (1992). *Human-computer interaction: psychology, task analysis and software engineering.* London: McGraw-Hall.

Johnson, P. (1998). *Usability and mobility.* Paper presented at the 1st Workshop on Human-Computer Interaction with Mobile Devices, Glasgow.

Kirakowski, J., & Corbett, M. (1988). Measuring user satisfaction. In D. M. Jones & R. Winder (Eds.), *People and computers IV* (pp. 329-430). Cambridge: Cambridge University Press.

Kirakowski, J., & Corbett, M. (1993). SUMI: the software usability measurement inventory. *British Journal of Educational Technology, 24,* 210-214.

Kjeldskov, J., & Stage, J. (2004). New techniques for usability evaluation of mobile systems. *International Journal of Human-Computer Studies, 60,* 599-620.

Knight, J. F., & Baber, C. (2005). A tool to assess the comfort of wearable computers. *Human Factors, 47,* 77-91.

Knight, J. F., Baber, C., Schwirtz, A., & Bristow, H. (2002). The comfort assessment of wearable computers. *Digest of Papers of the 6th International Symposium on Wearable Computing* (pp. 65-74). Los Alamitos, CA: IEEE Computer Society.

Monk, A. F., Wright, P. C., Haber, J., & Davenport, L. (1986). *Improving your human-computer interface: A practical technique.* London: Prentice-Hall.

Nielsen, J. (1993). *Usability engineering.* London: Academic Press.

Nilsson, J., Sokoler, T., Binder, T., & Wetcke, N. (2000). Beyond the control room: Mobile devices for spatially distributed interaction on industrial process plants. In P. Thomas & H. W. Gellerson (Eds.), *Handheld and ubiquitous computing: Second international symposium* (pp. 30-45). Berlin: Springer.

Nunes, L., & Recarte, M. A. (2002). Cognitive demands of hands-free phone conversation while driving. *Transportation Research Part F: Traffic Psychology and Behavior, 5,* 133-144.

Ravden, S. J., & Johnson, G. I. (1989). *Evaluating usability of human-computer interfaces.* Chichester: Ellis Horwood.

Slater, M., Usoh, M., & Steed, A. (1994). Depth of presence in virtual environments. *Presence: Teleoperators and Virtual Environments, 3,* 130-144.

Svenson, O., & Patten, C. J. D. (2005). Mobile phones and driving: A review of contemporary research. *Cognition, Technology and Work, 7,* 182-197.

Whiteside, J., Bennett, J., & Holtzblatt, K. (1988). Usability engineering: Our experience and evolution. In M. Helander (Ed.), *Handbook of human-computer interaction* (pp. 791-817). Amsterdam: Elsevier.

Wichansky, A. M. (2000). Usability testing in 2000 and beyond. *Ergonomics, 43*(7), 998-1006.

Wikman, A., Nieminen, T., & Summala, H. (1998). Driving experience and time sharing during in-car tasks on roads of different widths. *Ergonomics 4,* 358-372.

Wilson, J. R., & Nichols, S. C. (2002). Measurement in virtual environments: Another dimension to the objectivity/subjecitivity debate. *Ergonomics, 45,* 1031-1036.

Witmer, B. G., & Singer, M. J. (1998). Measuring presence in virtual environments: a presence questionnaire. *Presence: Teleoperators and Virtual Environments, 7,* 225-240.

KEY TERMS

Context of Use: The combination of user, task, product, and environment during the achievement of a desired goal

Dual-Task: The performance of two (or more) tasks at the same time. This could involve simultaneous performance or could involve some form of time-sharing between the tasks.

Effectiveness: The ability of a given user to employ a given product to achieve a desired goal in a given context of use

Efficiency: The optimal expenditure of resources by a given user in using a given product to achieve a desired goal in a given context of use

Referent Model: A product (or set of metrics) against which a given product can be compared

Satisfaction: The subjective response of a user to interacting with a product

Usability: *"… the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use."* (ISO9241-11, 1998)

ENDNOTES

[1] SUS: Software Usability Scale (Brooke, 1996)

[2] SUMI: Software Usability Metrics Inventory, Kirakowski and Corbett (1993)

Chapter XLIV
Usability Evaluation Methods for Mobile Applications

Regina Bernhaupt
University of Salzburg, ICT&S Center, Austria

Kristijan Mihalic
University of Salzburg, ICT&S Center, Austria

Marianna Obrist
University of Salzburg, ICT&S Center, Austria

ABSTRACT

Evaluating mobile applications and devices is particularly challenging given the variability of users, uses, and environments involved. This chapter introduces usability evaluation methods (UEMs) for mobile applications. Over the past decades various usability evaluation methods have been developed and implemented to improve and assure easy-to-use user interfaces and systems. Since most of the so-called 'classical' methods have demonstrated shortcomings when used in the field of mobile applications, they were broadened, varied, and changed to meet the demands of testing usability for mobile applications. This chapter presents a selection of these 'classical' methods and introduces some methodological variations for testing usability in the area of mobile devices and applications. It argues for a combination of both field evaluation methods and traditional laboratory testing to cover different phases in the user-centered design and development process.

INTRODUCTION

Usability testing of mobile applications is an emerging area of research in the field of human-computer interaction (HCI). It is commonly accepted that data collection for evaluation of mobile devices and applications is a central challenge, and that novel methods must be found for that (Isomursu, Kuutti, & Värinämo, 2004).

Overall, the study of the phenomena in the field of mobile HCI is highly driven by technology and concentrates primarily on producing solutions rather then reflecting on the methodologies in use. So far, only a few in-depth studies of the methodology utilized within the field of mobile HCI have been undertaken. Wynekoop and Conger (1990) describe the following methods in the face of mobile HCI: case-studies, field studies, action research, laboratory experiments, survey research, applied research, basic research, and normative writings. Kjeldskov and Graham (2003) conducted one of the most comprehensive reviews of the mobile HCI research methods. In their review, Kjeldskov and Graham (2003) selected papers focusing on mobile HCI from relevant conferences over the past years and classified them according to the described research methods. The findings clearly show that the prevalent mobile HCI research falls into the applied category (55%) followed by laboratory experiments (31%). On the bottom portion of the scale, no entries were found for action research, while only three and four (out of 102) research methods were conducted as case and field studies respectively. These findings imply that there is a strong tendency towards environment independent and artificial settings in research, while research on real use and action basic research is still widely neglected.

The evaluation of systems takes place predominantly in laboratory settings (e.g., Jones, Buchanan, & Thimbleby, 2002; Mizobuchi, Mori, Ren, & Michiaki, 2002). This lack of real-use-contexts and natural setting research could be justified by the fact that mobile HCI has strong roots in the field of computer sciences and HCI. These fields have a strong bias towards engineering and evaluation methodology. In addition, the methods stemming from the social sciences are normally more costly and resource intensive.

Recently, the need of addressing the question of what is useful and what is perceived problematic from a user perspective and under real use conditions has been noted. Increasingly researchers are incorporating natural setting methods—case and field studies and action research—either as standalone methods or in combination with artificial settings (e.g., Tamminen, Oulasvirta, Toiskallio, & Kankainen, 2003; Bohnenberger, Jameson, Krueger, & Butz et al., 2002; Hibono & Mockus, 2002). These methods offer ideal opportunities to study real-world user cases, to increase learning from existing systems within real-world contexts and assist with the translation of needs into new designs.

A good design, including usability and user friendliness, is not only critical for market success, but can also reduce mental and physical stress, reduce the learning curve, improve user-device operability when using the device. All these factors together improve the overall product quality (Duh, Tan, & Chen et al., 2006). As already mentioned, numerous methodologies are used for designing products and evaluating their usability. These methods are also applied to mobile device and application designs. The next section gives an overview of usability evaluation methods used.

USABILITY EVALUATION METHODS FOR MOBILE APPLICATIONS

The current trend of users demanding mobile information and communication technologies (ICTs) to support their everyday life and work has led to new generations of mobile devices. Mobile devices have expanded their functionality step by step. Looking at today's generation of mobile phones, various functions are offered. People may communicate via voice and text (short message service), receive information from the Internet, or use calendars on their cell phone to organize their daily lives. An endless variation of functionality exists on these pocket-sized devices. Mobile devices are used in various situations and contexts. They are used to support social life, peoples' social network, and are enablers of new ways of communication and coordination behaviors. The development of these mobile devices with their specific applications must follow an iterative process using appropriate usability evaluation methods at each stage of development to ensure the usability of mobile devices and applications.

Most of the usability evaluation methods developed during the last 25 years have focused primarily on desktop applications. The challenges for usability evaluation of mobile devices stem from their special characteristics like small screens with low resolution (compared to desktops), less power supply, and the trend to make devices smaller and smaller. At the same time the number of functions supported has increased. Non-standardized software development due to various operating systems used on these devices additionally complicate the matter. But these peculiarities are not the result of hardware and software trends only. Mobile devices are used in a variety of environments and contexts: at home, on the move, especially during travel. The location of the users is not the only determinant for the usage of mobile devices. The devices are used in a number of ways and situations not only influenced by the location, but rather by our activities: how we coordinate getting to places, how we adapt our daily routines, and how we organize and define our social networks are central usage behaviors for mobile devices. The context of use has a high impact on mobile device usage, and thus must be appropriately reflected in the usability evaluation.

Some of the methods developed for standard desktop applications have been adopted to be used during the development process of mobile devices and applications. A selection of some of the variations and adoptions of classical methods to fit usability evaluation of mobile devices is described in the following.

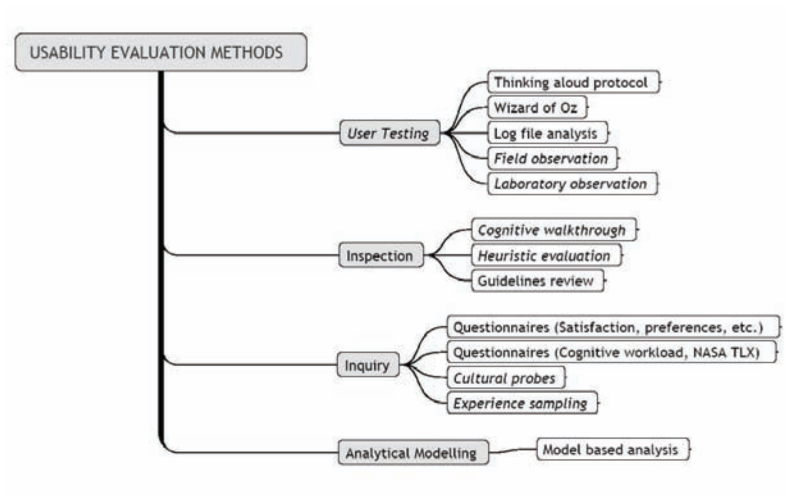
Framework for Current UEMs

Usability evaluation methods are classified within the following framework as:

- User testing (in the laboratory and the field)
- Inspection oriented methods (like heuristic evaluation and cognitive walkthrough)
- Self-reporting and inquiry oriented methods (like diaries and interviews)
- Analytical modeling (task model analysis and performance models)

This classification is based on traditional classifications for UEMs. More general classifications for methods in the field of mobile HCI are

Figure 1. An overview of currently used UEMs in the field of mobile HCI



available focusing on the way data is collected. For example, Hagen, Robertson, Kon, and Sadler (2005) distinguish between mediated data collection, simulations and enactments, and combinations of these two approaches for a more general framework on methods in mobile HCI. Which of the above evaluation methods is used depends on the strengths and weaknesses of the method, as well as its applicability with regard to the researcher's objectives (Duh et al., 2006).

User Testing

Usability tests are performance measurements to determine whether usability goals have been achieved. The usability evaluation of mobile devices in the lab has several advantages: the conditions for conducting the test can be controlled; all participants experience the same setting leading to higher quality data. On the other hand, user testing in the field allows discovery of usability problems related to the context of use. Pascoe, Ryan, and Morse (2000) studied the effects of using mobile devices while on the move, especially HCI related issues involved in using mobile devices in a real world (field) environment. The main factors which influence the interaction between users and mobile devices include contextual awareness, task hierarchy, visual attention, hand manipulation, and mobility. These are critical issues of usability and mobile application design. Testing in a real environment means that test users are able to behave as they would normally, to the extent that they need not to be told to do so.

Kjeldskov, Skol, Als, and Hoegh (2004) found that taking usability studies in the field only added little value, discovering the same problems both in the lab as well as in the field. On the other hand, several mobile HCI research studies assume benefits when conducting user testing in the field. Kjeldskov et al. (2004) argued that "… expensive time in the field should perhaps not be spent on usability evaluation (in the field) if it is possible to create a realistic laboratory setup including elements of context …" Duh et al. (2006) reported that more usability problems could be found in the field compared to the lab. To cope with the shortcomings of testing in the laboratory, several methodological variations and combinations of various methods have been proposed. For example, Lee, Hong, and Smith-Jackson (2006) propose the SEM-CPU approach—a systematic evaluation method for cell phone user interfaces, showing clearly the different kind of data that can be gathered during an experimental usability study.

Whether usability studies should be conducted in the lab or in the field is still a matter of discussion and needs further research. Conducting both kinds of studies has been argued for (see "Case Study: Towards a Real World Lab"), carefully selecting the necessary methodological approach based on the research question addressed.

How to conduct usability testing in "real-world" situations is also addressed in other areas of HCI. For example Bennett, Lindgaard, Tsuji, Conelly, and Siek (2006) conducted a workshop on how testing in non-traditional environments can be conducted. Experts from domains like military, healthcare, or mobile devices, discussed how methods have to be adopted to address testing in non-traditional environments. The general agreement was that only a mix of several methods during design and evaluation can help to ensure usability of the product.

Inspection

Inspection oriented UEMs are commonly used in industry because they are said to be fast and cheap. Most prominent is the usage of heuristic evaluation and cognitive and pluralistic walkthrough (Nielsen & Mack, 1994). It has been recognized that inspection-oriented methods lack validity when applied to mobile devices (Johnson, 1998; Kjeldskov & Graham, 2003). These methods do not take into account the contextual factors that affect user-system interaction. The success of these methods lies in the expert's ability to interpret the context of use and to draw meaningful conclusions.

To adopt the cognitive walkthrough method to reflect the context of use, Gabrielli, Mirabella, Kimani, and Catarsi (2005) enhanced the method using (cheap) video for performing the cognitive walkthrough in a typical environment. They explored the benefits of providing experts with more data about the user-system interaction. They found that providing evaluators with a video recording showing the usage of the system in a real environment or doing the evaluation in a real context increases the number of problems detected by 63 % and 68 % respectively, compared to conducting a normal heuristic evaluation. However, they also found that the number of problems discovered in the video enhanced situation compared to the real context situation, did not differ. They explain this fact with the expert's difficulties to note and use the method while moving and while being distracted or interrupted by external events. The number of problems found related to functionality was nearly the same for all three conditions (standard, video, real context), but the number of problems found related to use-context increased.

Po, Howard, Vetere, and Skov (2004) conducted a study to explore the interrelations between mobile use, heuristic evaluation, and the use context. They enhanced heuristic evaluation by two ways of contextual information. On the one hand, they extended heuristic evaluation with scenarios in use and conducted a so-called heuristic walkthrough in the lab. On the other hand, they varied the method by combining heuristic evaluation with scenarios in use and performing the method in the intended context of use. In this situation the scenarios as well as the intended context help evaluators by providing contextual cues. The heuristic evaluation and the two methodological variations were conducted by 11 experts in total. The heuristic evaluation discovered fewer usability problems than the two methodological adaptations. Po et al. (2004) also reports that heuristic evaluation did not discover the serious flaws with respect to the two developed variations. The standard heuristic evaluation showed no errors related to temporal sequences or time, while the varied methods found problems like that. Especially usage problems (like entering data in a lift) were discovered more frequently with the varied methods. The authors conclude that adding scenarios to heuristic evaluation helps to increase problem coverage, especially for more severe usability problems.

The modification of inspection oriented methods helps evaluators to combine the advantages of these methods and reduce the shortcomings of the methods when applied in the field of mobile systems and applications. These methods are still easy to conduct, fast and cheap compared to other approaches like experimental usability studies. They help to give quick feedback in an iterative development process. They can be used early in the development process. On the other hand, they still have the disadvantages of all expert oriented inspection methods, relying on the knowledge and expertise of the experts. This might be a reason why today the evaluation of mobile systems and devices in HCI oriented research is mainly done by (experimental) user studies, predominantly taking place in laboratory settings (e.g., Jones et al., 2002; Mizobuchi et al., 2002).

Inquiry (Adapted Methods)

To evaluate mobile devices and applications, questionnaires and self-reporting methods are additionally used to survey users' behaviors and usage of the systems. As traditional questionnaires conducted at the end of the usage are able to show several hindering effects (primacy and recency effects, etc.) new variations of these methods have been developed. Especially, different in-situ methods and how they help to address the context of mobile usage will be described in the next section.

There are many different methods available to perform in-situ user research. These methods include ethnography, (video) observations, contextual inquiries, diary studies, probe studies, or context mapping. Each approach has its own focus, advantages, and disadvantages. The methods used in earlier research like observations, contextual inquiries, interviews, and workshops are limited to a few work contexts, where the research environment was often highly organized by work structures (Kort & de Vries, 2006). These factors made the need for in-situ research methods less prominent since users can often easily recall their activities, needs, and preferences in detail, under these circumstances. With regard to the increasing amount of user research in mobile and private contexts, in-situ methods are becoming more and more important. It is impossible to observe users in their many different, especially mobile contexts, with limited resources and a limited time span for research. Moreover, users have difficulties recalling their activities, needs, and preferences since the private or mobile contexts are not as well structured as their work counterparts. Regarding these methodological challenges, new approaches have been adopted in the field of HCI in general and in mobile HCI in particular.

Using ethnography in the design has become increasingly prominent within HCI (e.g., Blomberg, Giacomi, Mosher, & Swenton-Wall, 1993; Simonsen, 1997; Crabtree, 1998). Ideally, an ethnographer does the field work over an extended period of time to obtain data (Fetterman, 1998). Due to the constraints of time and resources in user research, more practical approaches like rapid ethnography have been developed (Millen, 2000). In this approach, traditional ethnographic methods are combined with self-reporting methods and techniques. In the self-reporting approach, the ethnographer does not go into the field for a long time to collect the data. Instead, the researcher prepares tools that guide the participants in understanding how to collect the data themselves (Sanders, 2002). Cultural probes for example, provide a variety of ways to apply self-reporting in user studies and evaluation. In the following section, the cultural probes approach as well as the already mentioned experience sampling method (ESM), and how they have been considered for studying people's actions in mobile contexts, are discussed in more detail.

Cultural Probes

The cultural probes method was developed by Gaver, Dunne, and Pacenti (1999). The probe approach plays an important role in the initial phase of a user-centered design process. Cultural probes are purposefully designed to inspire, reveal, and capture the forces that shape an individual life in different contexts, like at home, at work, or on the move (Hemmings, Crabtree, Rodden, Clarke, & Rouncefield, et al., 2002).

Probes are mainly used to gather insights on the users' context in order to better inform the design process in an early stage (Gaver et al., 1999; Jääskö & Mattelmäki, 2003). Thereby, probe packages are provided to the study participants and consist normally of diaries, cameras, post cards, sometimes maps of the explored environments, and several other means, to obtain as much data as possible from the participants and his context. Apart from the traditional cultural probe approach, new methodological variations like domestic, technology, or mobile probes have been developed (Hutchinson, Mackay, Westerlund, Bederson, Druin, Plaisant, Beaudouin-Lafon, Conversy, Evans, Hansen, Roussel, Eiderbäck, Lindquist, & Sundblad, 2003; Crabtree, Hemmings, Rodden, Cheverst, Clarke, Dewsbury, Hughes, & Rouncefield, 2003; Arnold, 2004; Hulkko, Keinonen, Matttelmäki, & Virtanen, 2004). Mobile probes are mainly used to explore the mobile environment in order to explore people's activities in mobile contexts, but it is not a usability evaluation method.

Technology probes however, are more useful for evaluation purposes. Technology probes involve the deployment and evaluation of a new technology or application into a real use context (Hutchinson et al., 2003). "Technology probes are not primarily about usability in the HCI sense. They are not changed during the use period based on user feedback. In fact, a deliberate lack of certain functionality might be chosen in an effort to provoke the users." (Hutchinson et al., 2003). The main advantage of this approach for usability evaluation purposes is providing real life insights on a number of usability issues (e.g., how users choose to interact with a certain device in a special context), which have to be considered in the user-centered design and development process of new products and applications.

Cultural probes can inspire design by providing a wide range of material reflecting important aspects of the participant's context of use and potential usability factors by using technology probes. A more profound understanding of the mobile context provides a better basis for usability evaluations in a later stage in the design cycle. Technology probes particularly provide advancements to existing user testing evaluations. Recent shortcomings of laboratory user testing of mobile devices and applications can be overcome by using technology probes. Users can explore the technology in their real life setting without being directly observed by the evaluator. By using self-reporting materials (mainly diaries and photos) the user becomes the evaluator themselves over a specified period of time (e.g., 2 or more weeks). Probes results are normally discussed with the users in a post-interview and further explored within the design team in order to improve existing concepts and applications. One of the main challenges of probe studies related to mobile applications is to motivate people to complete the probe material in mobile contexts. In order to address this shortcoming, further methods are used, especially experience sampling.

User Experience Sampling

The experience sampling method is a technique from the field of psychology which addresses several evaluation needs for mobile and ubiquitous computing (Consolvo & Walker, 2003). "The Experience Sampling Method (ESM) is a research procedure for studying what people do, feel, and think during their daily lives" (Larson & Csikszentmihalyi, 1983). The ESM is conducted in-situ, involves many participants, and takes place over time, and collects quantitative and qualitative data. When using experience sampling for usability evaluation of mobile applications, the specific research interests as well as the measurement method, which are suitable to gain the desired information, must be carefully considered .

The main qualities of experience sampling are that usability and user experience factors can be studied within a natural setting, in real time, on repeated time occasions, and by request (Kort & de Vries, 2006). Computerized experience sampling on mobile devices has recently gained a lot of attention, especially since people are used to carrying mobile devices with them most of the time. Computerized experience sampling involves the use of mobile devices (like Palms or palmtop computers) on which the sampling software is installed and allows the participants to answer questions about their experiences in daily life. This method is getting easier and less expensive but it is still time and resource intense. Conner, Barrett, Bliss-Moreau, Lebo, and Kaschub (2003) provide additional information on ESM procedures and Van Esch-Bussemakers, Neerincx, Lindenberg, and Streefkerk (2005) report experiences about the combination of ESM with classical usability evaluation methods.

One of the main disadvantages of ESM is that it disrupts the user's activity, requiring the user to stop the current activity and answer questions on paper or on an electronic device. Therefore, Intille, Kukla, and Ma (2002) propose the use of image based experience sampling, where photos and short video clips are captured and can provide rich contextual information to the designer. An extension of this home-based approach combined with diverse logging and sensing measurements, can be considered for the mobile context. Future usability evaluations of mobile applications should also consider context-aware experience tools as well as the combination and triangulation of different in-situ methods (Intille, Rondoni, Kukla, Iacono, & Bao, 2003).

Analytical Modeling

In addition to the aforementioned user and expert centered usability evaluation methods and adopted methods for user inquiries, we should not forget traditional usability evaluation methods originally applied in the field of software engineering, like automatic testing, formal descriptions of user and task models, model-based evaluations, or critical incident and accident analysis. The authors are not aware of a broad usage of these methods in the field of mobile devices and applications, although one example in reference to Lee et al. (2006) was proposed in the SEM-CPU approach of a critical incident analysis.

Further methods to improve usability evaluation in mobile settings can be adopted from other areas and disciplines. For example, usability evaluation settings in the area of ubiquitous computing or smart homes can be successfully applied to address contextual issues. Koskela, Väänönen-Vainio-Mattila, and Lehti (2004) show how a smart home can help to address contextual issues while testing mobile applications, whereas Intille, Tapia, Rondoni, Beaudin, Kukla, Agrwal, Bao, and Larson (2003) describe how tools can support testing (of ubiquitous) applications in natural settings. This work inspired the setup for the development of a kind of "Real World" laboratory described in the following section.

Case Study: Towards a Real World Lab

Shortcomings of Traditional Usability Testing

Since mobile applications are commonly used en route, various stationary methods cannot be directly applied. Laboratory settings have the advantage of being able to control the environment (Love, 2005). Laboratory experiments are only partly feasible for contextual mobile applications. Such experiments enable testing of a system in an artificial, controlled environment. While they allow precise metrics, and thus more accurate comparisons and evaluations, the artificiality of the situation does not take external factors into account. However, when testing mobile applications, this might not be the preferred choice.

Mobile systems are usually used in different situations, which cannot be easily represented in the lab. Moreover, context issues are almost impossible to setup in a lab because of the different factors and their mutual interdependences. Except for classical usability issues (Kjeldskov et al., 2004) as well as testing some particular issues within a predefined context, testing of mobile applications and services might better be conducted in the field. Even mobile laboratories (sometimes called contextual laboratories), in which equipment is brought to the users' natural environment, cannot fully overcome these drawbacks. Ideally, in a contextual laboratory it would be possible to control different external factors in order to provide the same settings to all participants. Although field-testing does not provide the control over the settings, mobile applications are used in dynamic environments, thus it is important to assess these factors in-situ.

The Role of Context

Today's technology provides the users with the possibility to access information and communication services wherever they are and whenever they want. The social eco-system and the physical environment provide a vast amount of communication capabilities and information access, strongly influencing our everyday lives. Bearing this in mind, the context of interaction plays a major role in mobile settings. Understanding the particular context in which the systems are being used is crucial for the development of systems with good user experience and high user satisfaction.

People use mobile technologies such as mobile phones in a variety of settings: at work, in their homes, while on the streets, in public transport, while driving a car, in restaurants. Sometimes people use them also in theaters and lectures, or other quiet places where the usage of mobile phones can be annoying or even disrupting. The versatility of the technology catapulted the cell phone to become people's constant companion. This brings about an important issue: contrary to many other technologies, the context of use with mobile technologies is continuously changing. These changes are rapid and on a short-term, following the ever increasing mobility of the users themselves.

The concept of context is a problematic one. One reason for this is that there are so many aspects of context that depend on the current application and usage schemes. Thus it is difficult to come up with one general solution of evaluating contextual issues—taking a variety of context dimensions into account, each and every situation is entirely different from one another. However, it is exactly this variety of factors that constitutes context of our daily environment. The scope of contextual features is defined dynamically and is particular to each occasion of an action. Context is actively produced, maintained, and enacted in the course of the activity.

Traditional system design often fails to analyze and incorporate the social context in which the actions unfold. As Dourish (2004) argues, it is important to understand "how and why, in the course of their interactions, do people achieve and maintain a mutual understanding of the context for their actions?" Thus, methods are needed to assess and understand the interactional context of the users, especially in mobile settings and future ambient intelligence environments. In mobile environments, internal factors of the users, such as tasks and goals, are interconnected with external factors of the environment and the social system, which are dynamic and unpredictable (Blom, Chipchase, & Lehikoinen, 2005). Gathering data wherever the users are as well as adapting to the changes in the context is crucial in studying the use of mobile technologies.

The "Real World" Lab

The ICT&S Center for Advanced Studies and Research in Information and Communication Technologies and Society at the University of Salzburg seeks to enhance the positive potential of information and communication technologies for individuals, organizations, and society and to reduce negative phenomena. It incorporates three research units that work in a trans-disciplinary setting: HCI & usability, e-policy and e-society, as well as e-theory.

In order to cope with the aforementioned drawbacks of testing, a radical new laboratory has been operational since autumn 2005 at the ICT&S Center. Its concept is new insofar, as it allows different testing settings to be set up dynamically. On a standard scale it provides a typical lab setup with a controlling desk and a testing area with participant's PC (optionally using an eye-tracking device). This basic setup is already flexible enough to enable the testing area to be rearranged for testing situations other than PC-based interaction. Conducting tests of interactive TV services, remote controls, and mobile devices in a leisure room setting is accomplishable with minor effort. Likewise, providing a working environment for testing Web-based portals or collaborative software is easily achieved. The lab is also flexible enough to enable switching between setups during the tests, which is required for example in multi-channel or multi-modal applications.

On the next level, it is possible to move outside of the lab for conducting the tests. The ICT&S Center features two courtyards adjacent to the lab. This way the test moderator can be in the lab while the test participant can freely move around the courtyard. The inner courtyard is completely enclosed by the offices and the coffee lounge and can be monitored by the cameras. The outer courtyard is equipped with a night-vision camera and is directed to the street. Since the components of the lab are mobile, other areas of the ICT&S Center can be used for testing as well. The areas include a meeting room with a projection screen for testing work and collaborative settings, a lounge with a plasma screen for testing interactive technologies for leisure and spare time, and a multipurpose work area which features a wall-sized dual-head projection for testing 3D applications and games. Thus the entire ICT&S Center acts as a flexible user experience test-bed.

Expanding the limits even further, the nearby surroundings of the lab and the ICT&S Center provide the basis for conducting the tests "in the wild". A public square located just opposite of the outer courtyard enables testing under real conditions. It is partly covered by the outdoor camera and additional cameras can easily be positioned throughout the square. The cameras (and other testing equipment) can simply be hooked into the public wireless infrastructure which is available at the square and co-provided by the ICT&S Center. In the future, an extended setup is being prepared in collaboration with the municipality, in which an interactively arranged public space will allow interactive installations to be spread across the public square, and thus be used for usability and user experience experimentation.

The entire setup allows flexible usability and user experience observations beyond the limits of common usability study facilities—in particular moving out of the lab. In addition, the testing setup is also available as a mobile lab—albeit as a less flexible variation.

Figure 2. Real World Lab of the ICT&S Center of the University of Salzburg

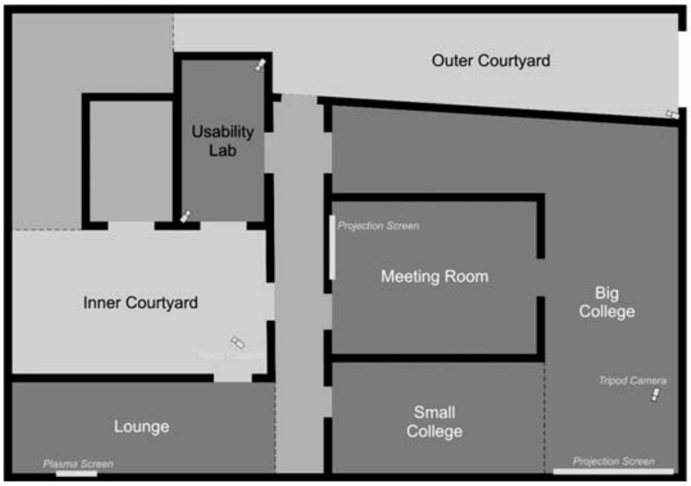


Figure 3. Different test settings for usability evaluation

The technology of the Usability Lab exploits the digital potentials based on a sophisticated network infrastructure. It incorporates a flexible eye-tracking system that can be used for different plane surfaces, because it is not integrated into the monitor. Network cameras with pan/tilt/zoom can be positioned wherever there is a wired or wireless IP-based network available, and are thus highly mobile. A mobile device camera allows capturing users' interaction with various mobile devices during the test in a wired or wireless setting. In addition, the setup integrates user experience measurement facilities, for example, bio-physiological measurements. Due to the digital nature of the lab, it is possible to make the testing sessions available worldwide and in real-time during testing. Thus, the clients are able to follow the test from anywhere in the world by the means of a Web browser. The flexibility of the lab enables the development of test-beds in various areas of interaction innovation—mobile interfaces, ambient intelligence, multi-modal interfaces, games, and so forth. Different types of sensors, gaze and body movement, display technologies, and several other technologies will be adapted and integrated to enable their adoption in several scenarios.

DISCUSSION AND OUTLOOK

Most important for usability evaluation is the selection of the appropriate methods in the various phases of the development process—to finally ensure usability aspects of the mobile devices and applications. Usability evaluation methods traditionally used for testing desktop applications have been adopted and improved to address the challenges in the field of mobile devices and applications. To enable data collection within usability evaluation, a set of methods are used, carefully selected based on the targets of the usability evaluation for the various devices and application. The aim is also to adopt and use methods from other fields in order to better inform the design process.

Based on the review of existing, evolved and adopted evaluation approaches, and methods for mobile devices and applications, it can be concluded that traditional UEMs are highly usable especially when they are adopted to reduce the shortcomings of the method (e.g., heuristic evaluation with scenarios). Even if there are different opinions and experiences with laboratory and field testing, there is a tendency to combine these approaches with regard to mobile application testing. Both methods have their advantages and disadvantages. A combination of field (in-situ) evaluation methods with traditional lab testing are recommended in order to cover different phases in the user centered design and development process.

From the experience in the "Real World" lab of the ICT&S center, it can already be asserted better that the spatial extension of the lab addresses the evaluation of mobile applications and devices. Further research will focus on how to test mobile applications in a test-bed outside the lab—in the real world. Additional techniques and tools (like self-reporting methods as probing) can support the evaluation in mobile contexts. Thus, forms of usability evaluation methods inspired by developments in the area of ubiquitous computing, enabling the monitoring of users in everyday life and during everyday usage of products, will be studied.

REFERENCES

Arnold, M. (2004). The connected home: probing the effects and affects of domesticated ICTs. In B. Adrian (Ed.), *Artful integration: Interweaving media, materials and practices* (Vol. 2) *Proceedings of the Eighth Biennial Participatory Design Conference (PDC'04).*

Bennett, G., Lindgaard, G., Tsuji, B., Connelly, K. H., & Siek, K. A. (2006). Reality testing: HCI challenge in non-traditional environments. *Workshop at CHI '06 extended abstracts on human factors in computing systems* (pp. 1679-1682). Montreal, Canada.

Blom, J., Chipchase, J., & Lehikoinen, J. (2005). Contextual and cultural challenges for user mobility research. *Communications of the ACM, 48*(7).

Blomberg, J., Giacomi, J., Mosher, A., & Swenton-Wall, P. (1993). Ethnographic field methods and their relation to design. In D. Dchuler, A. Namioka (Eds.), *Participatory design: Principles and practices.* Erlbaum: New Jersey.

Bohnenberger, T., Jameson, A., Krueger, A., & Butz, A. (2002). Location-aware shopping assistance: Evaluation of a decision-theoretic approach. In F. Paterno (Ed.), *Human computer interaction with mobile devices* (pp. 155-169). Berlin: Springer.

Conner, T., Barrett, L. F., Bliss-Moreau, E., Lebo, K., & Kaschub, C. (2003). A practical guide to experience-sampling procedures. *Journal of Happiness Studies, 4,* 53-78.

Consolvo, S., & Walker, M. (2003). Using the experience sampling method to evaluate ubicomp applications. *IEEE Pervasive Computing, 2*(2), 24-31.

Crabtree, A. (1998). Ethnography in participatory design. In *Proceedings of Participatory Design Conference* (PDC '98) (pp. 93-105). Palo Alto, CA: CPSR.

Crabtree, A., Hemmings, T., Rodden, T., Cheverst, K., Clarke, K., Dewsbury, G., et al. (2003). Designing with care: adapting cultural probes to inform design in sensitive settings. In *Proceedings of OzCHI'03: New Directions in Interaction, Information Environments, Media and Technology.*

Dourish, P. (2004). What we talk about when we talk about context. *Personal and Ubiquitous Computing, 8*(1),19-30.

Duh, H. B. L., Tan, G. C. B., & Chen, V. H. (2006). Usability evaluation for mobile devices: A comparison of laboratory and field tests. In *Proceedings of MobileHCI 2006* (pp. 181–186).

Fetterman, D. M. (1998). *Ethnography: step by step* (2nd ed.). Thousand Oaks: Sage Publications.

Gabrielli, S., Mirabella, V., Kimani, S., & Catarsi, T. (2005). Supporting cognitive walkthrough with video data: A mobile learning evaluation study. In *Proceedings of MobileHCI 2005* (pp. 77–82). ACM Press.

Gaver, B., Dunne, T., & Pacenti, E. (1999). Design: Cultural probes. *Interactions, 6*(1), 21-29.

Hagen, P., Robertson, R., Kan, M., & Sadler, K. (2005). Emerging research methods for mobile technology use. In *Proceedings of CHISIG* (pp. 1-10). Australia: ACM Press

Hemmings, T., Crabtree, A., Rodden, T., Clarke, K., & Rouncefield, M. (2002). Probing the probes. In *Proceedings of the Participatory Design Conference* (pp. 42-50). Palo Alto, CA.

Hibono, S., & Mockus, A. (2002). handiMessenger: Awareness-enhanced universal communication for mobile users. In F. Paterno (Ed.), *Human computer interaction with mobile devices* (pp. 170-183). Berlin: Springer.

Hulkko, S., Keinonen, T., Mattelmäki, T., & Virtanen, K. (2004). Mobile probes. In *Proceedings of Nordi CHI 2004* (pp. 43-51). ACM Press.

Hutchinson, H., Mackay, W., Westerlund, B., Bederson, B. B., Druin, A., Plaisant, C., Beaudouin-Lafon, M., Conversey, S., Evans, H., Hansen, H., Roussel, N., & Eiderbäck, B. (2003). Technology probes: Inspiring design for and with families. In *Proceedings of the CHI'03* (pp. 17-24). ACM Press.

Intille, S. S., Kukla, C., and Ma, X. (2002). Eliciting user preferences using image-based experience sampling and reflection. In *Proceedings of CHI'02, Short talk: User studies—Lessons for HCI* (pp. 738-739). Minnesota, USA.

Intille, S. S., Rondoni, J., Kukla, C., Iacono, I., & Bao, L. (2003). A context-aware experience sampling tool. In *Proceedings of CHI'03, Posters: Computer everywhere* (pp. 972- 973). Florida, USA.

Intille, S. S., Tapia. E. M., Rondoni, J., Beaudin, J. Kukla, C., Agrwal, S., Bao, L., & Larson, K. (2003). Tools for studying behavior and technology in natural settings. In *Proceedings of Ubiquitous Computing 2004* (pp. 157–174).

Isomursu, M., Kuutti, K., & Väinämö, S. (2004). Experience clip: Method for user participation and evaluation of mobile concepts. In *Proceedings of the Participatory Design Conference* (pp. 83-92). Toronto.

Jääskö, V., & Mattelmäki, T. (2003). Observing and probing. *In Proceedings of Designing pleasurable products and interfaces Conference (DPPI'03)* (pp. 126-131). Pittsburgh, USA.

Johnson, P. (1998). Usability and mobility: interaction on the move. In *Proceedings of the First Workshop on Human-Computer Interaction with Mobile Devices.* Glasgow, UK.

Jones, M., Buchanan, G., & Thimbleby, H. (2002). Sorting out searching on small screen devices. In F. Paterno (Ed.), *Human computer interaction with mobile devices* (pp. 81-94). Berlin: Springer..

Kort, J., & de Vries, A. (2006). ZAPP-ing user needs in situ: studying user needs in different contexts. In *Proceedings of the Workshop on Innovative Mobile Applications of Context MobileHCI'06.* Helsinki, Finland.

Kjeldskov, J., & Graham, C. (2003). A review of mobile HCI research methods. In L. Chittaro (Ed.), *Human-computer interaction with mobile devices and service* (pp. 317-335). Berlin: Springer..

Kjeldskov, J., Skov, M. B., Als, B. S., & Høegh, R. T. (2004, September 13-16). Is it worth the hassle? Exploring the added value of evaluating the usability of context-aware mobile systems in the field. In *Proceedings of Mobile Human-Computer Interaction—MobileHCI 2004: 6th International Symposium* (pp. 61–73). Glasgow, UK.

Koskela, T., Väänönen-Vainio-Mattila, K., & Lehti, L. (2004). Home is where your phone is: Usability evaluation of mobile phone UI for a smart home. In *Proceedings of the MobileHCI 2004* (pp. 74 – 85).

Larson, R., & Csikszentmihalyi, M. (1983). The experience sampling method. In H. T. Reis (Ed.), *Naturalistic approaches to studying social interaction* (pp. 41-56). San Francisco: Jossey-Bass.

Lee, Y. S., Hong, S. W., Smith-Jackson, T. L., Nussbaum, M. A., & Tomioka, K. (2005). Systematic evaluation methodology for cell phone user interfaces. *Interacting with Computers, 18*(2), 304-325.

Love, S. (2005). *Understanding mobile human-computer interaction.* Oxford/Burlington: Elsevier.

Millen, D. R. (2000). Rapid ethnography: Time deepening strategies for HCI field research. In *Proceedings of Designing Interactive Systems* (pp. 280-286).

Mizobuchi, S., Mori, K., Ren, X., & Michiaki, Y. (2002). An empirical study of the minimum required size and the minimum number of targets for pen input on the small display. In F. Paterno (Ed), *Human computer interaction with mobile devices* (pp. 184-194). Berlin: Springer.

Nielsen, J., & Mack, J. (1994). *Usability inspection methods.* NY: Wiley and Sons.

Pascoe, J., Ryan N., & Morse, D. (2000). Using while moving: Human-computer interaction issues in fieldwork environments. *Annual Conference Meeting Transactions on Computer-Human Interaction*, *7*(3), 417-437.

Po, S., Howard, S., Vetere, F., & Skov, M. B. (2004). Heuristic evaluation and mobile usability: Bridging the realism gap. In *Proceedings of MobileHCI 2004* (pp. 49–60).

Sanders, E. (2002). *Ethnography in NPD research. How applied ethnography can improve your NPD research process.* Retrieved October 8, 2006, from http://www.pdma.org/ visions/apr02/applied.html

Simonsen, J. (1997). Using ethnography in contextual design. *Communications of the ACM, 40*(7), 82-88.

Tamminen, S., Oulasvirta, A., Toiskallio, K., & Kankainen, A.(2003). Understanding mobile contexts. In L.Chittaro (Ed.), *Human-computer interaction with mobile devices and service* (pp. 17-31). Berlin: Springer.

Van Esch–Bussemakers, M. P., Neerincx, M. A., Lindenberg, J., & Streefkerk, J. W. (2005). User experience testing in context: Attentive services for the police. In *Proceedings of CHI2005 Workshop "Usage analysis: Combining logging and qualitative methods."* Portland, USA.

Wynekoop, J. L., & Conger, S. A.(1990). A Review of computer aided software engineering research methods. In H-E. Nissen, H. K. Klein, & R. Hirschheim (Eds.), *Information systems research: Contemporary approaches and emergent traditions* (pp. 129-154). Copenhagen: Elsevier.

KEY TERMS

Context: Mobile services and devices that can be used in various places and situations, by a single user or involving others. These circumstances are described as context of use or usage context.

Cultural Probes: An inquiry method aimed at inspiring, revealing, and capturing individual lives in different contexts, like at home, at work, or on the move.

Experience Sampling Method (ESM): An in-situ method especially suitable for collecting quantitative and qualitative data with mobile and ubiquitous systems. ESM studies user experience factors in a natural setting, in real time, and over a longer period of time.

Inspection-Oriented UEMs: A set of methods used by experts and most commonly based on guidelines to investigate possible usability problems.

In-Situ Evaluation Methods: A set of methods used to evaluate a system or prototype in its real usage context.

Usability Evaluation Methods (UEMs): A set of methods used to evaluate a system, mock-up, or prototype in terms of usability.

Usability Test: Performance measurements of users to determine whether usability goals have been achieved.

Chapter XLV
Evaluating Context-Aware Mobile Interfaces for Professionals

Jan Willem Streefkerk
TNO Defense, Security and Safety, The Netherlands
Delft University of Technology, The Netherlands

Myra P. van Esch-Bussemakers
TNO Defense, Security and Safety, The Netherlands

Mark A. Neerincx
Delft University of Technology, The Netherlands
TNO Defense, Security and Safety, The Netherlands

Rosemarijn Looije
TNO Defense, Security and Safety, The Netherlands

ABSTRACT

Evaluation refines and validates design solutions in order to establish adequate user experiences. For mobile user interfaces in dynamic and critical environments, user experiences can vary enormously, setting high requirements for evaluation. This chapter presents a framework for the selection, combination, and tuning of evaluation methods. It identifies seven evaluation constraints, that is, the development stage, the complexity of the design, the purpose, participants, setting, duration, and cost of evaluation, which influence the appropriateness of the method. Using a combination of methods in different settings (such as Wizard-of-Oz, game-based, and field evaluations) a concise, complete, and coherent set of user experience data can be gathered, such as performance, situation awareness, trust, and acceptance. Applying this framework to a case study on context-aware mobile interfaces for the police resulted in specific guidelines for selecting evaluation methods and succeeded to capture the mobile context and its relation to the user experience.

INTRODUCTION

In designing mobile support systems, evaluating designs at various stages in the development process is used to refine and adjust the design when needed. Furthermore, evaluation validates that the user needs and requirements are met for the intended user group. Thorough evaluations are required when the risks and costs of errors are high, when innovative interactive support systems, such as context-aware systems, are developed, or when the system is designed for use in a dynamic and critical environment. These needs for evaluation are even higher for mobile user interfaces, because of the dynamic use context, specific constraints of devices and risks of negative transfer from desktop experiences to mobile experiences (Nagata, 2006).

Due to these three issues, the user experience of mobile user interfaces is still an important bottleneck for services in the professional domain (Marcus & Gasperini, 2006). Realizing adequate user experiences is done by selecting the right method, based on specific constraints for evaluation of mobile, context-aware applications. Combining evaluation methods should capture the dynamic context aspects and their relations to the user experience in a complete, concise, and coherent way (cf. Neerincx & Lindenberg, in press). Finally, tuning of techniques and measures should ensure that the obtained results are relevant to the application domain.

Application Domain

The professional domain can be characterized as an environment where mobile workers are dependent on correct and relevant information to make critical decisions, where individuals are trained for their tasks and where tasks are goal-directed. In these domains, context-aware mobile devices have potential to support specific tasks such as notification to relevant information in context or facilitating communication with specific team members. Example domains include the police, ambulance and firefighter services, Urban Search and Rescue (USAR) teams, and the armed forces. Context-aware mobile devices have not yet become widespread in these domains.

Evaluation for the professional domain is distinguished from other domains by the following aspects. First of all, evaluation methods and measures should be tuned to specific user experience criteria within the application domain. For example, it seems less relevant (although interesting) to ask police officers about their emotional response toward the interaction with a mobile device. It seems more relevant to measure how many more criminal cases get solved in less time than before the introduction of the device. Secondly, not all situations for which the device is intended can be assessed in the field. Situations may not happen frequently enough or the risks are too high. For these situations, other research settings such as simulators may prove useful. Finally, access to professional end-users for evaluation purposes may be limited due to busy schedules and limited resources. The following case study is used to focus the discussion of evaluation methods for mobile context-aware interfaces and to provide an example from the professional application domain.

Case Study: Evaluating a Mobile Support System for Police Officers

For mobile police officers, increasingly more (multimedia) information becomes available to perform their tasks. In addition, both the interaction possibilities with devices and the momentary user needs for information or services continuously change over time and place (Baber, Haniff, Sharples, Boardman, & Price, 2001). Finally, shared situation awareness (SA) and communication within or between teams are vital for task execution, but may be diminished due to distributed persons and locations. Both theory and police practice show a clear need for interfaces that attract and guide the attention of individual officers or teams to relevant, high priority information or objects in a mobile setting (Streefkerk, Van Esch-Bussemakers, & Neerincx, 2006). The PAUI (Personal Attentive User Interface) project aims at designing and evaluating an adaptive user interface to support mobile police officers.

Throughout this chapter we will use the police officers' surveillance task as an example. On surveillance, police team members have to detect criminal incidents or respond to incoming calls and take fast and effective action. This requires notification of relevant information at the right time and place and optimal situation awareness, for instance, by knowing exactly where they are and where their colleagues are. A context-aware system can support these processes. In order to do this, the support system needs to have knowledge about the user, his use context, and the task he is working on. Subsequently, the system can adapt the interaction and communication to this knowledge. In this case study, a context-aware mobile system is designed that notifies police officers to incidents, based on their location and task. The main question here is to which context, user, and task factors this system should adapt and how it should adapt the interaction. Evaluation plays an important part in this project. At various moments, evaluating design solutions and concepts checks their validity for end-users, the use context, and the application domain.

Mobile Use Context

Context-aware mobile user interfaces are developed to improve the user experience by adapting the system behavior, based on a model of relevant use context factors. User experience is a term used to describe cognitive, affective, and social responses that are induced by the use of a product or service. However, the actual effect on user performance and acceptance has been assessed insufficiently for current applications (Goodman, Brewster, & Gray, 2004). Traditionally, evaluation is limited to laboratory settings and lacks the use of methods such as survey research, case study research, and evaluation in real use contexts that give validity to the research results (Kjeldskov & Graham, 2003). Use context is especially important for mobile devices as it can change constantly, in contrast with the use context of desktop applications. Although recreating central aspects of the mobile use context in the lab is sufficient to identify usability problems (Kjeldskov, Skov, Als, & Hoegh, 2004), the added value of field evaluation lies primarily in a deeper insight into the user experience in a dynamically changing context. In addition, field evaluation provides insights into effects of environmental factors, such as distractions, lighting conditions, body movement, and unreliable wireless networks (Duh, Tan, & Chen, 2005; Zhang & Adipat 2005).

The lack of field evaluation characterizes the professional domain as well. Only a few field evaluations of context-aware systems for professionals have been documented in the literature. In one effort to design context-aware support for firefighters, the application was evaluated with end-users outside the use context, although a field study was used to guide the initial design (Jiang, Chen, Hong, Wang, Takayama, & Landay, 2004). Results showed that the application was accepted by the firefighters and it supported their work practices. However, the researchers state that field testing of the application is necessary. A related project, Freeband FRUX, aims to design mobile applications for police and rescue workers (Van Eijk, De Koning, Steen, & Reitsma, 2006) by incorporating end-users in the analysis stage and a field test in the use environment. However, these projects are exceptions, stressing the need for a comprehensive approach to evaluation.

One of the problems with using real use contexts may lie in the fact that traditional evaluation methods are insufficient and inappropriate for evaluating context-aware applications in dynamic environments (Kellar, Inkpen, Dearman, Hawkey, Ha, et al., 2004; Vetere, Howard, Pedell, & Balbo, 2003; Zhang & Adipat, 2005). A shift can be seen towards employing new techniques to sample the user experience within the context of use. Examples are a heuristic walkthrough especially developed for mobile use (Vetere et al., 2003), and a context-aware questionnaire, which is presented to the user after a specific event. This results in more specific user reactions than using a general questionnaire (Kort & De Poot, 2005). However, these solutions are still in the development stage. Concluding from the discussion, evaluation of mobile context-aware systems is lacking a coherent and concise set of methods and techniques to "chart" the user experience in context. A more elaborate framework is necessary which takes into account the specific constraints of context-aware computing in the mobile, professional domain. This framework should provide guidance for the selection, combination, and tuning of evaluation methods. Furthermore, it should be flexible enough for evaluators who have different expertise and preferences. Finally, it should apply to other professional domains where mobile context-aware applications are designed and evaluated. This chapter proposes such a framework.

In the remainder of this chapter, first, the constraints in the evaluation of mobile adaptive systems for professionals are described. Then, following a user-centered design approach, the framework will be applied to the case study described. This framework is not intended to fully capture all existing evaluation methods, but to provide a practical approach for evaluation of professional mobile systems and present a "core" set of methods. Best practices, problems, and lessons learned are described in depth as they apply to the case study. It should be noted that this framework is general and can be applied to other evaluation methods than the ones mentioned in this chapter. Finally, specific guidelines for evaluation of mobile, adaptive systems are presented.

FRAMEWORK OF EVALUATION CONSTRAINTS

An effective and efficient use of evaluation methods is aimed at different moments to improve the quality of design solutions. However, selection of techniques is not straightforward as researchers are confronted with a diversity and multitude of evaluation methods and techniques. Kjeldskov and Graham (2003) propose a categorization of current mobile HCI research methods on the constraints of setting and purpose. They signal a lack of basic research and promote the development of theoretical frameworks to better describe, compare, and understand evaluation methods. Another framework for usability research methods for mobile devices is presented by Zhang and Adipat (2005). It emphasizes the setting of the evaluation (field vs. lab) based on the need to evaluate the application in context. While the frameworks help to select a particular research method, both lack specific guidance for deciding between and combining different evaluation techniques and measures in the evaluation of context-aware systems for professionals. Combination of methods should result in a more complete and sound knowledge base for design decisions, for example, by complementing and cross-validating results between methods. Further tuning of methods should ensure that results are relevant to the application domain. The framework distinguishes the following constraints that influence which methods, techniques, and measures can be employed.

Both the *stage* in the development process and the *purpose* of the evaluation set specific requirements for the available techniques. In addition, context-aware mobile systems are by nature, *complex*. The user interface changes due to the changing context and this emerging, adaptive behavior should be tested in a proper way. This complexity is increased by designing systems for the professional domain. The characteristics of this domain require a different approach than evaluating entertainment systems such as an MP3 player. Thus, the *setting* of the evaluation is important. Access to representative end-users and situations in the professional domain can be limited, begging the question *who* to include in your evaluation and for *how long.* Finally, different methods for evaluating mobile devices involve different *costs* in both time and resources. Concluding, seven constraints were identified that can be summarized as the following questions:

- Which stage of the development process are you currently in?
- What is the purpose of the evaluation?
- How complex is the design?
- Who are your participants?
- In which setting will the evaluation take place?
- What is the duration of the evaluation?
- What are the costs of the evaluation?

Stage in the Development Process

The development process for mobile context-aware applications can be separated into an analysis, design, and implementation stage. Mobile design solutions can be evaluated at every stage in the development process both within and outside of the actual use context.

The stage of the development process determines which techniques can be employed and what can be presented to participants during the evaluation. In early analysis stages, only high-level concepts and usage or problem scenarios are subject of an evaluation. In addition, the mobile work environment and tasks of professionals are analyzed, identifying tasks in need of support, problems in task execution, and appropriate characteristics to guide the context-awareness of the application. The focus is on gathering as much and diverse information as possible. In intermediate design stages, early versions of the adaptation model, mobile design solutions, and support for professionals' tasks can be evaluated on usability, appropriateness, and suitability for current work practices. Near the end of the process a functional demonstrator or prototype can be implemented. A benefit of early evaluation is that design flaws or errors are uncovered relatively early. Sometimes it suffices to evaluate only parts of a system, such as support for a specific task. Early prototyping and field testing is even more important for mobile applications than desktop applications as the usability of the mobile application is very dependent on the device used and the dynamic context (Zhang & Adipat, 2005). Here, evaluation provides an important proof of concept that the adaptation model and application result in meaningful support.

Purpose of Evaluation

A second constraint is the purpose of the study. For mobile, context-aware applications, purpose can be gathering factors on which to base adaptive system behavior, evaluate influence of environmental factors and mobility, or evaluate suitability for a specific task. It is distinguished between formative methods, used to generate design solutions, and summative methods, used to measure acceptance of designs. Within the framework, formative evaluation can be used to identify the factors on which to base the adaptive behavior. Contrastingly, summative evaluation focuses on how the system impacts the work processes of professionals and the correctness of the adaptivity model.

On a more fundamental level, the innovativeness of context-aware mobile systems also determines the purpose of evaluation. These evaluations must often take place without established benchmarks or design guidelines. In this case study, evaluation of revolutionary new concepts is adapted to specific police contexts and tasks. This purpose is in contrast with redesigning or improving existing applications.

Complexity of Design

How complex the design is constitutes the third constraint. Complexity in adaptive systems can be defined as "the directness of transformation from user input to system output" (Zipf & Jöst, 2005) that is, the adaptive system behavior. Design solutions with different degrees of complexity need different evaluation approaches. The evaluation of a calendar application on a mobile phone requires a different set of techniques and measures than the evaluation of a context-aware adaptive system. However, for mobile devices there should always be a fit to the dynamic context of use. This point is closely related to the innovativeness of the system. A factor that further increases the complexity is the fact that users themselves also show adaptive behavior. A system that dynamically adapts to dynamic user characteristics can cause unpredictable effects.

From a user perspective, evaluating adaptive systems means evaluating the appropriateness of the adaptive behavior, given the context and user task. Optimally, the system should be tested in the use context, because the adaptations are based on this use context. Depending on the goal of the evaluation, the question is whether or not to make the underlying rules or model explicit for users. Often the goal of an adaptive system is to seamlessly support the user's flow of work, making comparison to non-adaptive systems hard or irrelevant (Weibelzahl, 2005). In other situations, the adaptation rules or models need to be made explicit in order to be evaluated. Here, a "modular" approach could be adopted by evaluating the appropriateness of the input, the model, and the resulting behavior separately. This approach provides adequate feedback into the design process (Paramythis, Totter, & Stephanidis, 2001).

Participants

A fourth constraint is choosing the right participants and the right number of participants for testing. Evaluation shows to which extent the design meets the requirements of the end-user group. For professionals such as police officers, their diverse roles, skills, training, and experience impose specific requirements on the design (e.g., Pica & Sørensen, 2004). Determining these requirements, user characteristics and needs is the first step in evaluation. Next, during evaluation, an assessment is made how well the adaptive system supports specific roles or tasks. Often in professional settings, access to end-users is limited and deciding which method to use must take into account the availability of participants. End-users are particularly necessary during the analysis and implementation stages because of their knowledge of the mobile and dynamic use context and their work processes. In addition, prior training on or experience with certain tasks has to be taken into account, as well as prior experience with mobile devices. Negative transfer from desktop experience to mobile experience can cause longer task execution times and more switching between tasks (Nagata, 2006). When no actual end-users can be involved, a careful selection of participants has to ensure they are representative of end-users.

Setting of Evaluation

Furthermore, the setting of the evaluation is of importance. The setting of mobile systems evaluation can be defined as environment independent, natural, or artificial (Kjeldskov & Graham, 2003). Environment independent methods are not situated in the use environment. Their focus is on creating a general overview of system use instead of describing specific tasks. For context-aware mobile systems, gathering information about the use context is particularly important during the analysis stage. Hence, contrasting to evaluation of desktop applications, environment independent methods must be combined with methods that provide a rich description of the dynamic use context. The results can be captured in, for example, scenarios, storyboards, and use cases.

This contrasts with the natural or artificial setting of task-based evaluations. In essence, choosing between a natural or artificial setting is balancing a trade-off between the degree of reality of the evaluation setting and control over extraneous variables. The purpose of evaluation in a natural setting is proving that the system works as intended in a realistic use environment. For example, for context-aware systems, the correctness of the adaptive behavior with respect to the context is evaluated. However, when a high degree of control over extraneous variables is needed, an artificial laboratory setting can be used. Recreating or simulating essential elements from the use environment in the lab has specific benefits for evaluating professional systems. In this domain, field evaluation may interfere with ongoing work and imposes on the time of participants. In addition, situations for which the design is intended may not happen frequently enough to evaluate them properly, for example, large-scale disasters. In this case, a good alternative is to simulate the use environment and test the context-aware system in the lab (Te Brake, De Greef, Lindenberg, Rypkema, & Smets, 2006). Finally, if actual mobile use is subject of evaluation, simulation of an application on a real mobile device has advantages over simulation on a desktop computer. Specific constraints for the device and environmental factors (such as low bandwidth) are taken into account during the evaluation, providing more realistic results (Zhang & Adipat, 2005).

Duration of Evaluation

The duration of the evaluation is constrained by the type of data that is collected during evaluation. Some data can be collected relatively fast and easy by interacting with a prototype for a couple of hours. Examples include usability questionnaires or task performance data on a specific task. This data is focused and specific, that is, only valid for the task and can not be generalized to other tasks and settings. In contrast, evaluation in a longitudinal study gives deeper insight into how learning effects, the dynamics of trust, and user experience develop over time. These measures are particularly important in evaluating mobile, adaptive systems. Interpretation of this general, broad data makes it necessary to take into account the whole context of use (Kort & De Poot, 2005). Tuning measures to the application domain can be done by relating them to performance criteria for professionals (cf. Neerincx & Lindenberg, in press; see Box 1).

Evaluation Cost

Finally, the cost of an evaluation can be expressed in time and resources. Thus, the cost-effectiveness of the evaluation method can be viewed as the amount and severity of uncovered design flaws versus the cost of investing time and resources. For evaluating mobile applications, video logging with behavior analysis is a widely used but time consuming and expensive method of which the added value remains debated (Kjeldskov, Graham, Pedell, Vetere, Howard, Balbo, & Davies, 2005). Recent comparisons between methods show that rapid reflection by experts is a very cost effective procedure, uncovering the majority of critical usability problems in a short time. However, for evaluating mobile adaptive systems, issues like ecological validity of the design can only be tested in field situations. These studies entail higher costs due to the mobility of the setup and the participation of professional end-users. Furthermore, there is less room in the professional domain for flawed designs leading to usage errors, calling for a more extensive evaluation. Possible cost-efficient solutions for evaluating these innovative systems are using Wizard of Oz prototypes or simulations.

Conclusion

From the discussion, it is clear that the stage, purpose, complexity, participants, setting, duration, and cost each impose constraints on which evaluation technique to use. When the constraints to the goal of evaluating context-aware professional user interfaces are specifically applied, the following can be concluded:

- Evaluation within and outside the use context with participation of end-users can take place at every *stage* in the development process, each stage having its own focus.
- The *purpose* of evaluation is influenced by the innovativeness of the system and determines whether formative or summative techniques are used.
- Evaluating *complex* adaptive systems in the use context increases appropriateness of the final design.
- Actual end-users must be involved as *participants* because of their intimate knowledge of mobile use context and domain-specific tasks.
- Information about the dynamic use context must be gathered as early as possible. Furthermore, when access to the actual use *setting* is restricted, simulation yields a realistic yet controlled evaluation environment.
- Evaluation over longer periods of time in the mobile application domain is particularly important to gather rich, broad user experience data.
- Using simulation tools can reduce the *cost* of evaluation, but the user experience and ecological validity can only be evaluated in relatively expensive field testing.

It is important to note that all seven constraints are interdependent. For example, the setting of an evaluation depends on the participants as it makes little sense to evaluate a support for a specific police task using students in an artificial setting.

However, each constraint has its own unique contribution to the selection of techniques.

APPLYING THE FRAMEWORK

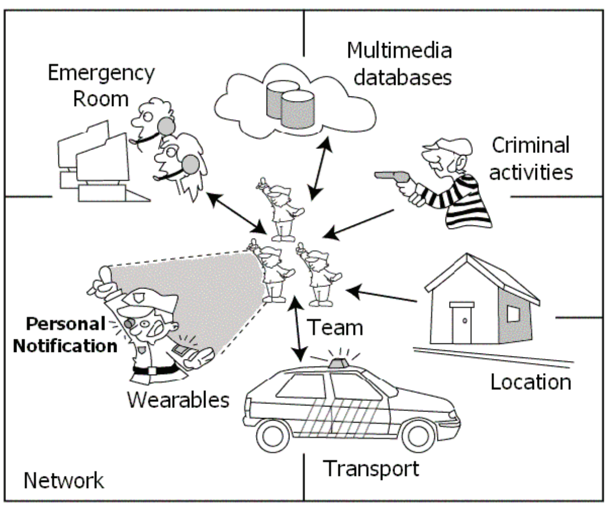
The framework of constraints to the case study of the PAUI project is now applied (described in the first section) on designing a context-aware notification user interface for police officers. In this case study, a user-centered design approach was followed, described in Box 1. Based on this approach, it is illustrated how the evaluation methods and techniques at each step in the UE method were selected, combined, and tuned and their benefits and limitations are discussed. The advanced stage of the project enables all methods to be addressed. However, it must be noted that this research is still ongoing and not all results are obtained yet. Theoretical considerations and relevant alternative techniques are presented in separate boxes (Box 2, 3, & 4). These techniques are not unique to evaluation of mobile context-aware applications. The boxes are categorized on the constraint of evaluation setting (environment independent, artificial, and natural).

The PAUI project started off with a definition of the concept. This concept was based on literature research of relevant HCI literature and domain research including participatory observation of the work domain. The concept for a support system for mobile police officers is shown in Figure 1. Police officers operate within a network of different information sources, such as the emergency room and (multimedia) databases. Information on criminal activities and their location is presented to a team of police officers, each having their own characteristics, task, and transport. Based on these factors, the individual officer will receive personal notification at the right time and place, possibly via a wearable or handheld device.

Focus Group Evaluation

- **Stage:** Early analysis stage
- **Purpose:** Innovative design; formative
- **Complexity:** High; concept of adaptive system
- **Participants:** Thirty; police personnel with diverse backgrounds and experience
- **Setting:** Within and outside use environment
- **Duration:** Short; one half day
- **Costs:** High in resources; low in time

Figure 1. Concept of the attentive user interface for police officers



Box 1. User-centered design method

The Usability Engineering (UE) method provides an empirical design and evaluation approach where knowledge about user needs and design solutions are refined until they meet the usability requirements. Because these cannot be predicted from the onset of design, involving users in the design process is regarded as an important necessity for a successful design. This method has been developed and applied for the design of user-interface support for space missions, ship control centers, and mobile services.

Approach

The UE method starts with the definition of a concept (see Figure 2, left side), which is a broad description of the proposed system. Scenarios are then drafted from the relevant application domain and describe users, their tasks, and context in a comprehensive, narrative style. Especially describing and understanding the dynamic nature of use contexts for mobile devices is crucial for applying UE. From the scenarios, the process of requirements analysis results in a requirements specification. These requirements describe in detail the user needs with respect to their work practice and the role the system fulfills in addressing these needs. User requirements form the basis for the system features. Features can be considered solutions to user needs and describe what functionality the system should have. As the method progresses from concept to features, the level of detail increases.

Evaluation

Evaluation of concept, scenarios, user requirements, and features is done by validating them to objective and subjective quality criteria, such as established Human Computer Interaction (HCI) metrics and new HCI metrics specifically adapted for mobile devices (see Figure 2, right side). Effectiveness, efficiency, and satisfaction are established criteria for evaluating HCI (ISO 9241-11, 1998) but should be matched to domain-specific performance criteria. Furthermore, professionals have to be able to rely on a system while performing critical tasks. Therefore, trust and user acceptance are important aspects that influences actual use and develop over time (Marsh & Meech, 2000). Finally, mobile context-aware systems should maintain or heighten situation awareness (SA) by providing the right information at the right time (Endsley & Garland, 2000).
It is important to note that this UE method is an iterative process, with a full cycle including the assessment of the proposed features on HCI metrics, and further specification of these features based on this assessment. Parts of the system can also be evaluated, it is not necessary to evaluate the whole system at once. The end products of this cycle are generic guidelines, models and prototypes that are validated in their respective context and use domain.

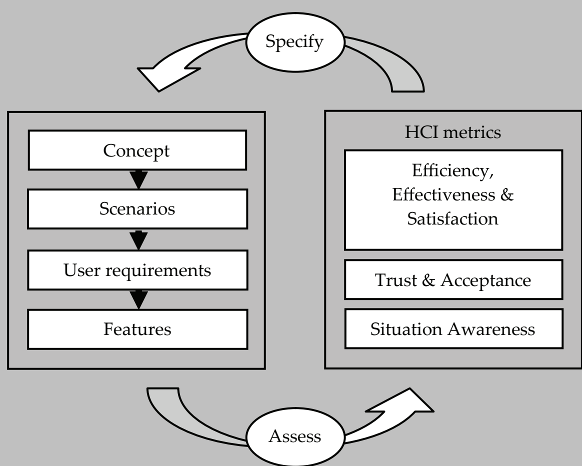
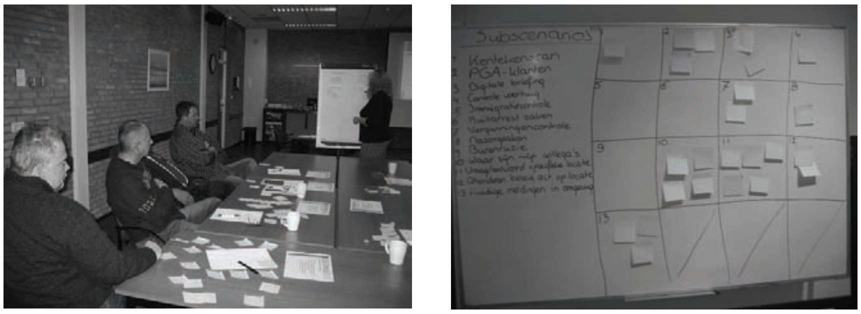


Figure 2. The Usability Engineering Method (adapted from Streefkerk et al., 2006)

Further Reading

- Neerincx and Lindenberg (2005)
- Lindenberg, Nagata and Neerincx (2003)
- Gorlenko and Merrick (2003)
- Vetere et al. (2003)

Figure 3. Illustrations from the focus groups showing the participants (left) and the plenary voting system (right)



Selection and Application of Method

To evaluate the high-level concept, involvement of end-users to capture the mobile police context and domain knowledge was needed. Furthermore, a brainstorm discussion of innovative use scenarios and context of use from different perspectives was wanted because the purpose was to gather as much and diverse information as possible. Finally, no concrete previous examples were available that could serve as a reference point. Based on these considerations, a focus group setting with police officers and management personnel was selected. Alternative methods such as questionnaires or survey research (see Box 2) did not satisfy these requirements.

During this focus group, the concept presented was explained and the participants brainstormed about possible situations where the context-aware system would have an added benefit (see Figure 3). In four small groups (separate for police officers and management personnel) of five to six participants, ideas were written down in the form of short stories with illustrations. The participants were stimulated to “think out of the box” and to give room for new and refreshing ideas. The groups were moderated by a researcher who guided and stimulated the discussion with examples and recorded comments. At the end of the session, a plenary vote was taken on which stories illustrated maximum benefit for police work practice and would be used further. Every participant could give one positive and one negative judgment and the totals were added up for every story. The scenario with the highest positive score was selected for further use.

The resulting scenario clearly illustrated problems police officers experience when on surveillance. Examples include quickly assessing when a call is relevant to them and what priority a call has. The participants considered these problems to result in unnecessary distraction from their surveillance task and to hamper optimal task execution. They also believed the concept to be a possible solution to these problems.

Combination of Techniques

To validate the results of the focus group, they were combined with rapid ethnography (see Box 4), conducted on surveillance with three police officers. This ethnography focused on the work context, communication tools, location information, and support for procedures that officers use. The results of the study identified relevant moments in the work of police officers that could use support from a context-aware system. For example, sometimes a call is only relevant during a specific shift or for a specific group of officers. Based on these moments, relevant task and context aspects (such as location, task priority, and officer task history) were identified. By charting similarities and consistencies between the focus group scenario and the results from the ethnography study, the validity of these aspects was supported by actual work practice.

In addition, another focus group was used to evaluate the resulting usage scenario from the first focus group. This second focus group was similar in setting and participants to the first. Its purpose was more restricted because participants were instructed to think of realistic work situations where a context-aware system could provide benefits. From the scenario and discussion with participants, a list of requirements resulted. Again, a plenary vote decided which requirements was either indispensable, necessary, or merely worth considering (see Figure 3). This technique resulted in a prioritized list of domain-specific requirements, validated by end-users.

Tuning of Techniques

Tuning of the focus groups and ethnography to the application domain was done in four ways: First, by involving a police officer in the organization and execution of the focus group. Second, by specifically focusing on problems that police officers on the move would encounter and could be solved by a context-aware system. For example, a context-aware system presents only calls that are in the direct vicinity of the police officer. Third, by relating the benefits of a context-aware system explicitly to performance criteria for police officers, such as amount of time spent on surveillance or response times to calls. Finally, during the focus group, by dividing the officers and management into two separate groups. This resulted in separate usage scenarios for officers and management and allowed relating scenario elements to specific user characteristics and roles.

Benefits and Limitations

This evaluation resulted in a usage scenario and list of requirements, validated by end users and domain analysis. By analyzing and observing the surveillance task of the police officers in context, relevant moments for notification were identified. In addition, relevant context aspects (such as time,

Box 2. Environment independent evaluation techniques

Focus group evaluation
For designing mobile, context aware systems, this evaluation method is best used at an early stage of the process, when user requirements need to be defined for the system. During a focus group session, a small, selected group of people is brought together for an interactive and spontaneous discussion on a specific topic. The purpose of a focus group session is to gather broad information and to get insight into user needs and opinions through interaction between group members. Focus group research can be used for evaluating concepts, scenarios, and high-level user requirements.

Interview
An alternative technique is interviewing domain-experts or expert users such as police end-users. Questions are asked to get expert opinions and deeper understanding of the problems in the domain. For designing in the professional domain, interviews with police end-users can help establish domain-specific evaluation criteria to which a context-aware system can be judged.

Survey
Distributing questionnaires to the end-user population is a way of getting a large quantity of opinions from a diverse group. Specific advantages of survey research to evaluation in the professional domain are that surveys provide an overview of the police organization and allow user needs to be related to specific roles. For example, a police officer on surveillance may have different needs than a police officer visiting a crime scene.

Further Reading

- Jiang et al. (2004)
- Kjeldskov and Graham (2003)

location, and history) were captured. Employing end-users helped to determine priorities for the requirements. The combination of focus groups and rapid ethnography resulted in a unique and validated list of requirements for the design of mobile, context-aware applications for the police. Alternative approaches to evaluating high-level concepts are discussed in Box 2.

Some limitations to the quality of the focus groups were observed. The participants found it sometimes difficult to relate the scenario to specific, realistic work situations and had selective recall for some situations. They focused quickly on established procedures and found it hard to integrate possibilities offered by new technologies. In addition, for the moderators, it was difficult at times, to interpret insights into the working practice of police officers.

Designing the Context-Aware System

After the second focus group, a list of requirements for the attentive user interface for police officers was compiled (Streefkerk et al., 2006). From this list, the requirement of notification was the first and most important focus. The attentive user interface should notify police officers to relevant information in their environment, without distracting them unnecessarily from their primary surveillance task. This was done by adapting the salience and information density of the notification (i.e., the notification style) based on two rules. First, when user workload was high, information was presented more concisely. For example, a short summary of the message was presented, prior to the actual message. Second, when message priority was high, the salience of the audiovisual signals was increased, for example, by using loud auditory signals. This adaptive notification principle was evaluated by simulating it on a handheld computer (PDA). The simulated system presented messages to the user in the different notification styles, based on their workload and the message priority (see Figure 4). High priority messages were presented with red visual flashing bars or icons and a sharp sound. Medium priority messages were presented with a soft sound and low priority messages without sound. In low workload situations the full message text was presented at once, but in high workload situations, first a summary of the message was presented.

Wizard of Oz Evaluation

- **Stage:** Intermediate design stage
- **Purpose:** Validate innovative design solution; formative
- **Complexity:** Moderate; (simulated) adaptive functionality
- **Participants:** Twenty; representatives

Figure 4. Screenshots from the PDA with a summary (left) and a high priority message (right)

- **Setting:** Artificial; lab experiment employing Wizard of Oz setup
- **Duration:** Short; 2 hours
- **Costs:** Low in both time and resources

Selection and Application of Method

To evaluate the innovative support concept of adaptive notification, a Wizard of Oz setup was chosen (see Box 3) based on the following considerations. The purpose of the evaluation was to guide the further design effort. A simulated setting that allowed recreated basic aspects of the police officers surveillance task was needed. Furthermore, a flexible environment was necessary because the influence of changing context (e.g., workload and message priority) on interaction with a mobile device was being tested. Finally, the Wizard of Oz setup allowed the concept to be empirically tested by systematically comparing two conditions. As the concept dealt with general instead of task-specific abilities, a representative participant group was used.

Twenty participants were involved in this study, representative to end-users in age and education. They had to perform a simulated police surveillance by watching videos, recalling targets, and answering questions on these videos (see Figure 5). Simultaneously, the researcher sent low, medium, or high priority messages at predefined moments to the PDA. Participants had to recognize and report the messages. Adaptive notification (different notification styles) was directly compared with non-adaptive notification (uniform notification styles) in a within-subjects design. Each evaluation took approximately 2 hours, including training, two scenarios, and debriefing.

Combining and Tuning of Techniques

This evaluation combined both qualitative and quantitative techniques. Performance data (time on task, number of errors in task, and questions) were collected using event-logging on the PDA and questionnaires. Subjective judgments (notification intrusiveness, preference for condition) were measured with rating scales and questionnaires. The specific performance measures were tuned to realistic aspects of the police officers' surveillance task. For example, the messages were representative of police reports. In addition, participants had to recall and describe different "targets" from the videos, which is an important surveillance skill. Results from the evaluation indicated that the adaptive notification is positively evaluated. Participants preferred the adaptive notification over non-adaptive and a trend towards better performance with the adaptive system was observed.

Figure 5. Screenshots from the videos used in the Wizard-of-Oz evaluation

Box 3. Evaluation in artificial settings

Wizard of Oz
The Wizard of Oz (WoZ) evaluation method is widely used in evaluation of mobile context-aware applications. It involves letting participants interact with a seemingly functional system (possibly in the mobile context) that is actually operated by the researcher. This avoids programming a functional context-aware system and allows for early and relatively low-cost evaluation of design solutions. However, the weakness of the WoZ technique is human intervention. This technique is appropriate when no time-critical system performance is required.

Game based evaluation
Game-based evaluation provides best of both worlds for evaluation of mobile applications: a realistic task environment with control over extraneous variables. It provides an ideal simulation environment for task-based evaluation for professionals. Control over context factors means that the application can be evaluated under a wide variety of situations. Measurement of performance data can be done accurately due to integrated logging procedures. In addition, data gathering tools do not have to be taken into the field to evaluate mobile technology. Game-based techniques have been used frequently in learning and training environments and as simulation for crisis management situations.

Further Reading

- Dahlback, Jonsson, and Ahrenberg (1993)
- Lewis and Jacobson (2002)

Benefits and Limitations

The Wizard of Oz setup managed to create the illusion of a working, adaptive support concept. Participants indicated they could compare the conditions easily, which improved accuracy and validity of their subjective judgments. However, additional training is necessary to facilitate the distinction between notification styles. This study delivered new insights into the user experience with an adaptive, context-aware system. It resulted in specific, validated notification styles, matched to user workload and message priority. Based on this evaluation, the concept is redesigned to employ more distinctive sounds and visual signals. Finally, it was found that the notification style has to match the task that has to be performed.

Game-Based Evaluation

- **Stage:** Intermediate design stage
- **Purpose:** Innovative; summative; validate adaptive notification for teams
- **Complexity:** High; (simulated) advanced functionality
- **Participants:** Twenty-four participants in teams of three will be included, depending on availability
- **Setting:** Artificial; lab experiment employing game-based environment
- **Duration:** Extended; 4 hours
- **Cost:** Low in both time and resources

Selection and Application of Method

In the previous Wizard of Oz study, participants could only make a limited set of decisions in a task they did not directly control. The next planned evaluation aims to evaluate the redesigned adaptive notification principle and model in a richer yet controlled environment. Game-based evaluation allows flexibility in recreating task-specific aspects of the use context, such as team tasks with multiple actors. The simulated reality of the task environment requires using end-users as participants. Furthermore, it allows measuring performance and shared situation awareness by accurately logging participants' behavior. Based on these considerations, a lab experiment in a game-based simulation environment was selected (see Box 3).

A surveillance environment will be created within the PC game Unreal Tournament (see Figure 6; for a description, see Te Brake et al., 2006). Including 24 participants in teams of three is aimed at, depending on availability. The team navigates through this environment on surveillance, including reconnaissance, gathering information, and communicating with team members. In addition, participants receive assignments for additional tasks (finding locations or items) via the context-aware system, simulated on the PDA. By modeling user workload, location, and task, it decides which participant to present with which task. Both performance and the appropriateness of the adaptive behavior are subjects of evaluation. Therefore, an experimental condition with the adaptive system will be compared to a non-adaptive system. The duration of the evaluation is approximately 4 hours to allow thorough training on using the environment.

Combining and Tuning of Techniques

During this evaluation a combination of qualitative and quantitative measures is collected. Performance data include time on task, number of errors, and distance traveled. In addition, a measure for effectiveness of the system would be the number of tasks solved. Trust, acceptance, and preference were measured using questionnaires and rating scales. Situation awareness (SA) is measured with a technique called "freezing" (Endsley & Garland, 2000) where the workflow is paused at irregular intervals to answer a question about the environment, such as "indicate on the map the location of the car accident." In addition, the "critical incidents" technique uses a think-aloud protocol to collect both positive and negative incidents in using the context-aware system.

The evaluation setup is tuned to the police environment by using a diverse set of tasks that are representative of police surveillance. Furthermore, the critical incidents reported by the police officers participating in this evaluation are analyzed carefully. These incidents may suggest the appropriateness of the context-aware system in the field.

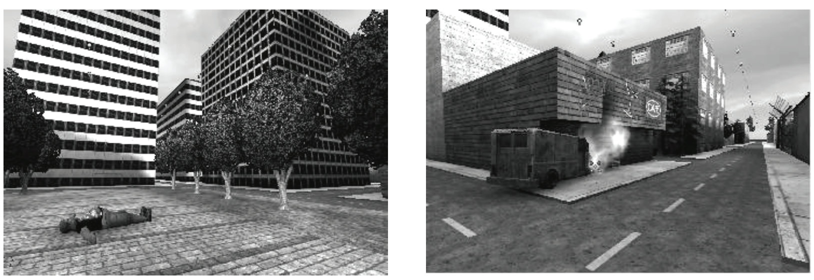
This evaluation is expected to result in a validation of the adaptive notification principle in a rich task-based setting. Furthermore, appropriate moments support is necessary, based on context factors such as task and location, are expected to be identified.

Benefits and Limitations

The game-based technique allows for accurate quantitative measures of performance data and SA, because the behavior and navigation path of the participant are recorded. In addition, the appropriateness of the adaptive behavior can be measured as well. Furthermore, multiple participants can work collaboratively on one task in the same environment, allowing evaluation with teams. Finally, critical events can be pre-programmed into the scenario running in the game simulation.

Some factors negatively influence a game-based evaluation. Asking participants to fill out

Figure 6. Screenshots from the game-based environment Unreal Tournament, showing a victim (left) and a car accident (right)



SA questions and rating scales interferes with the task flow at certain moments. In addition, prior gaming experience should be well documented, as this influences participants' performance. Finally, some participants are susceptible to simulator sickness, which can occur in game-based simulation (Kolasinski, 1996).

Field Evaluation in the Professional Domain

- **Stage:** Final implementation stage
- **Purpose:** Summative; validate functioning of final system in context
- **Complexity:** High; full system functionality
- **Participants:** Thirty; end-users
- **Setting:** Natural use environment
- **Duration:** Longitudinal; 3 months
- **Cost:** High in both time and resources

Selection and Application of Method

The final evaluation planned in the PAUI project will be field evaluation of the adaptive notification system in the natural work setting (see Box 4). The purpose is to validate the full functioning of the innovative context-aware system with end-users and to provide the final "proof of concept" in the application domain. This system is evaluated in a longitudinal study to measure impact on work processes, trust, acceptance, and learning effects. The costs in both time and resources are relatively high compared to other methods, as the police

Box 4. Evaluation in natural settings

Field evaluation is conducted in natural environments, often during the final phase in the development cycle and over longer periods of time. It requires a stable and reliable functioning system, participation of end-users, and mobile data gathering tools. As the functioning of the mobile system is dependent on the dynamic context and unreliable wireless networks, evaluating context-aware support systems in the field provides validation that the design works as intended. The added benefit of field evaluation over other methods has been criticized and disadvantages are possible interference with ongoing work, difficulties to encompass the richness of mobile contexts, and the difficult data collection and control due to the dynamic context and physically moving users.

Ethnography
To study mobile applications use through (rapid) ethnography, researchers immerse themselves in the work practice. They meticulously describe the context and common practices of the domain. A benefit of the technique is deeper insight into end-user practices in their natural work setting. This insight is of extra importance to understand the dynamic context of mobile end-users. Studying the police work environment provides a detailed description of common and uncommon tasks and critical incidents that a context-aware system can support.

Ethnographic field studies and field experiments
Within natural setting evaluation, a distinction can be made between ethnographic field studies and field experiments. An ethnographic study would describe the functioning of the mobile context-aware system in the work context and require participation of the researcher in the work activities. Contrastingly, field experiments would test two versions of a context-aware system under different conditions to evaluate the influence on task performance. Field experimentation allows for more control but can only be used for restricted evaluation purposes, such as a usability evaluation.

Further Reading

- Goodman et al. (2004)
- Kjeldskov et al. (2004)
- Zhang and Adipat (2005)

organization and personnel have to participate. In addition, collecting and analyzing field data is necessary, further increasing the costs.

The final prototype of the context-aware user interface will be implemented on a mobile device, integrated with existing police infrastructures and made available to police officers. The context-aware system notifies police officers to relevant information based on their location, provides location of colleagues, and supports task switching and police procedures. Thirty officers with different roles, such as emergency aid, district surveillance, and prevention, participate in this evaluation. After an initial training phase, the system is used during daily work for a period of 3 months.

Combining and Tuning of Techniques

This evaluation focuses on the user experience in context, integration of the system in work practice, and acceptance within the organization. Techniques include participatory observation, interviews, and questionnaires. To evaluate the user experience, critical incidents in task execution with the system are reported weekly by the officers. These reports are then related to the specific context variables logged by the system. Finally, researchers conduct a monthly participatory observation session on surveillance with officers. This technique aims at getting deeper insight into the system's impact on work processes.

The system is evaluated in a pre and post-test setup, thereby giving insight into changes caused by the system. Prior to evaluation, the expected effects of the system are captured in specific criteria. These performance criteria are tuned to the police application domain: the amount of fines collected, response time to calls, and amount of time spent on surveillance are important measures. This data is collected by recording events from police databases and analyzing system events on the PDA.

Benefits and Limitations

Observing professional end-users interacting with the system in their work environment gives insight into usability, user experience, and impact on work processes. In addition, only in field studies can the system be assessed in the actual and diverse work situations that occur naturally. This is a necessary and valuable step before actual implementation of the finished system, as it allows final changes and tuning of the system.

However, it is also a costly method as an advanced prototype, a mobile evaluation setup and the participation of end-users are necessary. In addition, at this implementation stage, it is difficult to make thorough changes in design when needed. Finally, judgments by participants (i.e., reporting critical incidents) should be given as fast as possible to avoid recall problems, but this may interfere with ongoing work.

EVALUATION GUIDELINES

In the third section, four specific methods were described to evaluate mobile context-aware technology for professionals. Based on the benefits and limitations of these techniques and the lessons learned in the PAUI project, the key findings are summarized in the form of guidelines. In Table 1 the appropriateness of the focus group, Wizard of Oz, game-based and field evaluation with respect to the seven evaluation constraints is presented. A plus sign (+) means the technique is appropriate considering that particular constraint. A minus sign (-) means the technique is less appropriate.

As can be concluded from the preceding discussion, evaluation of mobile, context-aware applications differs from desktop evaluation. Therefore, guidelines specific for evaluation of mobile, context-aware applications are printed bold in Table 1. For example, game-based evaluation techniques are less appropriate or necessary in desktop evaluation. The focus on the use context is more important in early stages for mobile

Table 1. Guidelines for the appropriateness of evaluation techniques (vertical) based on the constraints (horizontal)

		Focus group	Wizard of Oz	Game-based	Field
Stage	Analysis	+	-	- -	++
	Design	+	+	++	++
	Implementation	-	-	++	+
Purpose	Formative	+	+	++	+
	Summative	-	-	++	++
Complexity	Low	+	-	- -	-
	Medium	-	+	++	-
	High	+	-	++	+
Participants	Representatives	-	+	++	-
	End-users	++	++	++	+
Setting	Independent	+	-	- -	-
	Natural	-	++	- -	++
	Artificial	-	+	++	-
Duration	Short	+	+	++	+
	Longitudinal	-	-	- -	++
Costs	Time	+	+	++	-
	Resources	-	+	++	-

applications. In addition, to measure ecological validity in summative approaches, field testing is necessary. The participation of end-users with specific domain and context knowledge is more important than in desktop evaluation. In this way, the table can be used to get a quick overview of the appropriateness of these specific evaluation techniques.

FUTURE TRENDS

Two important future trends in evaluation of mobile, context-aware systems include employing more and diverse game-based techniques, user experience sampling in context, and mixed reality techniques. The increasing use of game-based evaluation techniques are easily explained by their advantages over other techniques (see Box 3) such as adaptability, flexibility, and accuracy in measurements, particularly for use for professionals such as police officers and rescue workers (Te Brake et al., 2006).

By employing creative solutions, limitations of traditional evaluation techniques can be reduced. For example, giving subjective judgments in rating scales out of the use context is difficult due to recall problems. By employing context-aware questionnaires that are triggered by specific system events or specific context factors (such as location), these recall problems are circumvented. An additional benefit is that the factors that triggered the questionnaire can be simultaneously logged, providing a deeper insight into the use context (Kort & De Poot, 2005). Another novel development is evaluation in "augmented" mixed reality settings. The participant is wearing a head-mounted display while performing a task in a natural setting. This display shows both the real environment and a layer on which extra information can be presented. By employing this setup, the power of simulation can be employed in natural evaluation settings.

CONCLUSION

Earlier research identified the lack of structuring in methods and techniques for evaluation of mobile and adaptive technology (e.g., Paramythis et al., 2001). The effects of this lack include inability to interpret and generalize results across applications and user groups. Traditionally, the focus in designing mobile systems has been on producing engineering solutions, rather than conducting ecologically valid evaluations, leading to a prevalence of lab evaluations (Kjeldskov & Graham, 2003).

In this chapter, it is argued that evaluation of mobile context-aware systems for professionals benefits from a systematic approach. As there is not one evaluation technique that delivers answers to all design questions, combinations of techniques have to be sought. By considering the development stage, the design complexity, the purpose, participants, setting, duration, and cost of evaluation, a specific set of methods, techniques, and measures can be determined. This framework of evaluation constraints was applied to a case study in designing support for mobile police officers. This resulted in specific guidelines for evaluation of mobile, adaptive systems for four specific techniques (focus group, Wizard of Oz, game-based, and field evaluation).

The approach helped to select a concise and coherent set of appropriate evaluation methods and techniques and to tune these to the appropriate application domain as was demonstrated by the case study. In addition, the framework contributed to the field of evaluation research by stating specific guidelines that reach beyond the current application domain and are suited for use in other domains as well. Evaluation of adaptive mobile systems in the professional domain is expected to specifically benefit from this approach. It stresses the need to incorporate end-users in the evaluation, emphasizes the critical and dynamic professional environment, and interprets evaluation results within the task and use context. In this way, both short-term usable services as well as long-term innovative support concepts for police officers are realized.

ACKNOWLEDGMENT

The PAUI project is part of MultimediaN and funded by the Dutch Ministry of Economic Affairs.

REFERENCES

Baber, C., Haniff, D., Sharples, M., Boardman, M., & Price, A. (2001). A requirements analysis of personal mobile computers for police officers. In D. Harris (Ed.), *Engineering psychology and cognitive ergonomics*, (Volume 6, pp. 21-28), *Industrial ergonomics, HCI, and applied cognitive psychology*. Aldershot: Ashgate.

Dahlback, N., Jonsson, A., & Ahrenberg, L. (1993). Wizard of Oz studies: Why and how. In *Proceedings of the 1st International Conference on Intelligent User Interfaces* (pp. 193-200). ACM Press.

Dix, A., Finlay, J., Abowd, G., & Beale, R. (1998). *Human-computer interaction* (2nd ed.). New York: Prentice Hall.

Duh, H. B. L., Tan, G. C. B., & Chen, V. H. (2006). Usability evaluation for mobile device: A comparison of laboratory and field tests. In *Proceedings of Mobile HCI'06* (pp. 181-186). Helsinki, Finland.

Endsley, M. R., & Garland, D. J. (Eds.). (2000). *Situation awareness: Analysis and measurement*. Mahwah, NJ: LEA.

Goodman, J., Brewster, S., & Gray, P. (2004). Using field experiments to evaluate mobile guides. In *Proceedings of the 3rd Annual Workshop on HCI in Mobile Guides*. Glasgow, Scotland.

Gorlenko, L., & Merrick, R. (2003). No wires attached: Usability challenges in the connected mobile world. *IBM Systems Journal, 42*(4), 639-651.

ISO 9241-11. (1998). *Ergonomic requirements for office work with visual display terminals (VDTs)—Part 11: Guidance on usability* (Tech. Rep. No. 9241-11).

Jiang, X., Chen, N. Y., Hong, J. I., Wang, K., Takayama, L., & Landay, J. A. (2004). SIREN: Context-aware computing for firefighting. In *Proceedings of Pervasive 2004*, Vienna.

Kellar, M., Inkpen, K., Dearman, D., Hawkey, K., Ha, V., MacInnes, J., Mackay, B., Nunes, M., Parker, K., Reilly, D., Rodgers, M., & Whalen, T. (2004). *Evaluation of mobile collaboration: Learning from our mistakes* (Tech. Rep. CS-2004-13). Canada: Dalhousie University, Faculty of Computer Science.

Kort, J., & De Poot, H. (2005). Usage analysis: Combining logging and qualitative methods. In *Proceedings of CHI '05 extended abstracts on human factors in computing systems*, (pp. 2121-2122).

Kjeldskov, J., & Graham, C. (2003). A review of mobile HCI research methods. In L. Chittaro (Ed.), *Proceedings of the 5th International Mobile HCI 2003 Conference* (pp. 317-335). Germany: Springer-Verlag.

Kjeldskov, J., Graham, C., Pedell, S., Vetere, F., Howard, S., Balbo, S., & Davies, J. (2005). Evaluating the usability of a mobile guide: The influence of location, participants and resources. *Behaviour and Information Technology, 24*(1), 51–65.

Kjeldskov, J., Skov, M. B., Als, B. S., & Hoegh, R. T. (2004). Is it worth the hassle? Exploring the added value of evaluating the usability of context-aware mobile systems in the field. In *Proceedings of Mobile HCI 2004*, Glasgow, Scotland.

Kolasinski, E. M. (1996). *Prediction of simulator sickness in a virtual environment.* Unpublished doctoral dissertation, University of Central Florida, Orlando.

Lewis, M., & Jacobson, J. (2002). Game engines in scientific research: Introduction. *Communications of the ACM, 45*, 27-31.

Lindenberg, J., Nagata, S. F., & Neerincx, M. A. (2003). Personal assistant for online services: Addressing human factors. In D. Harris, V. Duffy, M. Smith, & C. Stephanides (Eds.), *Human-centred computing: Cognitive, social and ergonomic aspects* (pp. 497-501). London: LEA.

Marcus, A., & Gasperini, J. (2006). Almost dead on arrival: A case study of non-user-centered design for a police emergency-response system. *Interactions, 13*(5), 12-18.

Marsh, S., & Meech, J. (2000). Trust in design. In *Proceedings of the 2000 Conference on Computer Human Interaction* (pp. 45-46). The Netherlands: The Hague.

Nagata, S.F. (2006). *User assistance for multitasking with interruptions on a mobile device*. Unpublished doctoral dissertation.

Neerincx, M. A., & Lindenberg, J. (2005). Situated cognitive engineering for complex task environments. In *Proceedings of 7th International Conference on Naturalistic Decision Making.* The Netherlands, Amsterdam.

Neerincx, M. A., & Lindenberg, J. (in press). Situated cognitive engineering for complex task environments. In J. M. C. Schraagen, L. Militello, T. Ormerod, & R. Lipshitz (Eds.), *Natural decision making & macrocognition.* Aldershot, UK: Ashgate Publishing Limited.

Neerincx, M. A., Ruijsendaal, M., & Wolff, M. (2001). Usability engineering guide for integrated operation Support in space station payloads. *International Journal of Cognitive Ergonomics, 5*(3), 187-198.

Paramythis, A., Totter, A., & Stephanidis, C. (2001). A modular approach to the evaluation of adaptive user interfaces. In *Proceedings of Workshop at the 8th International Conference on User Modeling, UM 2001* (pp. 9-24).

Pica, D., & Sørensen, C. (2004). On mobile technology in context: Exploring police work. *Journal of Computing and Information Technology, 12*(4), 287-295.

Streefkerk, J. W., Van Esch-Bussemakers, M. P., & Neerincx, M. A. (2006). Designing personal attentive user interfaces in the mobile public safety domain. *Computers in Human Behavior*, *22*, 749-770.

Te Brake, G., de Greef, T., Lindenberg, J., Rypkema, J., & Smets, N. (2006). Developing adaptive user interfaces using a game-based simulation environment. In: B. Van de Walle & M. Turoff (Eds.), *Proceedings of the 3rd International ISCRAM Conference*. Newark, NJ.

Van Eijk, R., De Koning, N., Steen, M., & Reitsma, E. (2006). *Developing a we-centric service for mobile police officers to support spontaneous communication*. (Tech. Rep. Freeband FRUX D1.8a). Available at https://doc.freeband.nl/dscgi/ds.py/Get/File-67631 (14-01-2007).

Vetere, F., Howard, S., Pedell, S., & Balbo, S. (2003). Walking through mobile use: Novel heuristics and their application. In *Proceedings of OzCHI* (pp. 24-32). Brisbane, Australia.

Weibelzahl, S. (2005). Problems and pitfalls in the evaluation of adaptive systems. In S. Chen & G. Magoulas (Eds.), *Adaptable and adaptive hypermedia systems* (pp. 285-299). Hershey, PA: IRM Press.

Zhang, D., & Adipat, B. (2005). Challenges, methodologies, and issues in the usability testing of mobile applications. *International Journal of Human-Computer Interaction, 18*(3), 293-308.

Zipf, A., & Jöst, M. (2005). Implementing adaptive mobile GI services based on ontologies—Examples for pedestrian navigation support. In *CEUS-Computers, Environment and Urban Systems—An International Journal* (Special Issue on LBS and UbiGIS).

KEY TERMS

Adaptive User Interface: A computer interface that adapts the interaction guided by a model of a specific set of factors

Context Aware Technology: Technology that models a set of factors about the use context to adapt the interaction with the user

Empirical Evaluation: An evaluation method in which results are derived by observation or experiment instead of theory

Ethnography: An evaluation method in the analysis stage during which end-users are observed and interviewed in their work environment

Game-Based Evaluation: An evaluation method in which a simulated environment is used adapted from existing computer games

Requirements Analysis: An analysis method during the analysis stage in which the user needs and requirements for design solutions are specified

User-Centered Design: A design methodology in which end-users needs and requirements guide the design choices

User Experience: The set of cognitive, affective, and social responses that are induced by the use of a product or service

Wizard of Oz Evaluation: An evaluation method in which a functioning system is simulated by a person controlling a non-functioning prototype

Chapter XLVI
Appropriating Heuristic Evaluation Methods for Mobile Computing

Enrico Bertini
DIUF - University of Fribourg, Switzerland

Tiziana Catarci
University of Rome "La Sapienza", Italy

Alan Dix
Lancaster University, UK

Silvia Gabrielli
University of Udine, Italy

Stephen Kimani
Jomo Kenyatta University of Agriculture and Technology, Kenya

Giuseppe Santucci
University of Rome "La Sapienza", Italy

ABSTRACT

Heuristic evaluation has proven popular for desktop and Web interfaces, both in practical design and as a research topic. Compared to full user studies, heuristic evaluation can be highly cost-effective, allowing a large proportion of usability flaws to be detected ahead of full development with limited resource investment. Mobile computing shares many usability issues with more conventional interfaces. However, it also poses particular problems for usability evaluation related to aspects such as limited screen real estate, intermittent user attention, and contextual factors. This chapter describes a modified collection *of usability heuristics that are designed to be appropriate for evaluation in mobile computing. They have been systematically derived from extensive literature and empirically validated. They therefore offer a sound basis for heuristic-based evaluation in mobile computing. Besides introducing the reader to the practical use of heuristic evaluation, the chapter also closes with a description of potential future research in the area.*

INTRODUCTION

Expert-based evaluation techniques, such as heuristic inspection (Nielsen et al., 1990) and cognitive walkthrough (Wharton et al., 1994) typically benefit from providing evaluators with guidance (for instance, a set of heuristics or a checklist) for identifying a prioritized list of usability flaws (Kjeldskov et al., 2005). Expert-based evaluation techniques are also well-known methods that can realize a relatively quick and easy evaluation.

According to Po (2003), mobile computing devices are typically 'smart products' or 'information appliances', and are generally consumer products. Their users are thus a 'heterogeneous group' (Sade et al., 2002) and so it may be more difficult to find suitable surrogate users for user-based testing in mobile computing (Po, 2003). Po further observes that even if appropriate surrogate users were found, the realistic recreation of the user context in laboratories would be challenging because of user mobility, which makes observation and video recording difficult (Vetere et al., 2003). "Given the problems associated with user-based evaluations of mobile devices, expert-based usability techniques are considered to be more appropriate" (Po, 2003). However, it is worth noting that expert-based techniques have in the past, been criticized for finding proportionately fewer problems in total and disproportionately more cosmetic problems (Karat et al., 1992). In mobile computing, the capacity of expert-based techniques to adequately capture the multiple contextual factors that affect user–system interactions in real settings has been questioned (for instance: Kjeldskov et al., 2003; Johnson, 1998).

It is believed that heuristic evaluation can be enriched and adapted toward capturing contextual factors. This chapter describes how standard heuristic evaluation can be made more appropriate/relevant for mobile computing. In particular, the chapter describes a modified collection of usability heuristics that are designed to be appropriate for this area. The heuristics have been systematically derived from extensive literature and empirically validated, and so offer a sound basis for heuristic-based evaluation of mobile computing. As well as introducing the reader to the practical use of heuristic evaluation, the chapter also describes potential future research in the area.

This work has been carried out in the context of MAIS[1], a project whose research goal is to provide a flexible environment to adapt the interaction and the information and services provided, according to ever changing requirements, execution contexts, and user needs.

The rest of the chapter is organized as follows: The second section highlights some of the challenges posed by mobile devices, applications, and context; the third section discusses the standard heuristic evaluation method; the fourth section describes the methodology that was adopted to appropriate heuristic evaluation for mobile computing and the results that were obtained; the fifth section contains reflections regarding this research activity; and the sixth section concludes the chapter and highlights some future work.

The focus in this chapter is on usability problems in mobile devices and a discussion of their sources. However, this should be set against a broader view of the fantastic world of new opportunities, advantages, and benefits that mobile devices and contexts bring. While it will not be explicitly touched on again, the chapter should be read in the light that the problems and limitations are ones worth tackling because of the opportunities offered by the technology.

MOBILE DEVICES, APPLICATIONS, AND THEIR CONTEXT

In order to better understand how usability in mobile computing can be evaluated and improved, it is useful to outline specific limitations inherent in mobile devices, applications, and their context. These fall into two broad categories: limitations due to the nature of the devices themselves and limitations due to context of use and style of interaction.

Limits Posed by the Device

- **Small-screen:** In order to be portable, mobiles must necessarily be small, and tend to have small screens, therefore problems due to the screen real estate are intrinsic and can be addressed only by figuring out new techniques to organize information visually.
- **Limited input:** Because of device format, input mechanisms are inherently limited. Currently, the most common means of interaction are: numeric keypads, which are used in almost all cell phones and styluses, which are the primary input means in PDAs and smart phones.
- **Limited bandwidth and cost:** Mobile Internet connections are still slow. This is in fact still one of the main factors limiting mobile Internet access. To this, the problem of the cost model must also be added. Most companies offer their Internet access in a pay per KByte policy that obviously limits the size of pages and the number of requests.
- **Limited connectivity:** Perhaps more than bandwidth, the latency of the connection affects its usability. The limited coverage of different networks and the consequent intermittent connection makes the latency extremely variable, as well as giving rise to problems of how to portray these hidden network properties to the user. There is also the problem of seamlessly switching between different types of network, for example, WiFi to GPRS.
- **Limited computational resources:** This means the capabilities of applications are limited. However, this should be overcome in the near future as new processors and memories specifically designed for mobile devices increase their quality and speed.
- **Limited power (batteries):** This is often an underestimated issue, but the batteries are still a big problem for every kind of mobile system (laptops included). This has a big impact on end users: limited autonomy means limited availability, which in turn means limited reliability.
- **Wide heterogeneity (of OSs and physical properties):** Users of mobile systems must always adapt to new forms of interaction as they switch to different mobiles. Changing the physical device and operating system usually translates into the need to re-learn functions, operations, messages, and so forth, with an enormous waste of resources.

Limits Posed by Context and Interaction

- **Variable context:** Since mobile devices, by definition, are mobile, the context in which they are used is continually changing. This poses challenging new issues because, though context has always been considered a fundamental aspect to analyse in usability studies, only now must such frequent and complex context variations within the same device, application, or single user, be addressed.
- **Kind of interaction:** The nature of interaction also changes in mobile settings. In general, users tend to interact in small and focused chunks of activities, more so than in fixed settings. A high proportion of tasks in a mobile environment consist of a few fast steps that the user should be able to execute without cognitive effort. In addition, mobile tasks may happen in conditions where users' attention is necessarily reduced, or may be part of more complex activities with which they should not interfere.

- **Interruptions:** Mobile devices/applications are always "with us." If this, on one hand, means that computation and data are always available, it is also true that notifications and requests for attention can happen at inappropriate moments and that some tasks may be interrupted. This raises two kinds of problems: appropriateness of notifications and recovery from interruptions.
- **Privacy and security:** Privacy issues become more prominent. While staying mobile, users find themselves in a variety of spaces (private and public), in a variety of situations (formal and informal), and in a variety of infrastructures (wireless and cable connection). Moving through these settings means having different needs for privacy and security.
- **Intimacy and availability:** Because mobile devices are mobile, they are personally available in a way that fixed devices are not. Moreover, they seem to engender a sense of being "personal" in a deeper sense than desktop PCs (e.g., not just my PC but my PDA and definitely my phone).

HEURISTIC EVALUATION

Introduction to Heuristic Evaluation

Heuristic evaluation (Nielsen et al., 1990; Nielsen, 1994b) is an inspection usability evaluation method. In heuristic evaluation, experts scrutinize the interface and its elements against established design rules. The experts should have some background knowledge or experience in HCI design and usability evaluation. Three to five experts are considered to be sufficient to detect most of the usability problems. The enlisted experts individually evaluate the system/prototype under consideration. They assess the user interface as a whole and also the individual user interface elements. The assessment is performed with reference to some usability heuristics. When all the experts are through with the assessment, they come together and compare and appropriately aggregate their findings. In Molich et al. (1990) and Nielsen et al. (1990) Rolf Molich and Jakob Nielsen initially proposed a set of usability heuristics for the design of user interfaces. Aiming to maximize the explanatory power of the heuristics, Nielsen later refined them (Nielsen, 1994b), thereby deriving the following set:

1. **Visibility of system status:** The system should always keep users informed about what is going on, through appropriate feedback within reasonable time.
2. **Match between system and the real world:** The system should speak the users' language, with words, phrases, and concepts familiar to the user, rather than system-oriented terms. Follow real-world conventions, making information appear in a natural and logical order.
3. **User control and freedom:** Users often choose system functions by mistake and will need a clearly marked "emergency exit" to leave the unwanted state without having to go through an extended dialogue. Support undo and redo.
4. **Consistency and standards:** Users should not have to wonder whether different words, situations, or actions mean the same thing. Follow platform conventions.
5. **Error prevention:** Even better than good error messages is a careful design which prevents a problem from occurring in the first place. Either eliminate error-prone conditions or check for them and present users with a confirmation option before they commit to the action.
6. **Recognition rather than recall:** Make objects, actions, and options visible. The user should not have to remember information from one part of the dialogue to another. Instructions for use of the system should be visible or easily retrievable whenever appropriate.
7. **Flexibility and efficiency of use:** Accelerators—unseen by the novice user—may often speed up the interaction for the expert user such that the system can cater to both

inexperienced and experienced users. Allow users to tailor frequent actions.

8. **Aesthetic and minimalist design:** Dialogues should not contain information which is irrelevant or rarely needed. Every extra unit of information in a dialogue competes with the relevant units of information and diminishes their relative visibility.
9. **Help users recognize, diagnose, and recover from errors:** Error messages should be expressed in plain language (no codes), precisely indicate the problem and constructively suggest a solution.
10. **Help and documentation:** Even though it is better if the system can be used without documentation, it may be necessary to provide help and documentation. Any such information should be easy to search, focused on the user's task, list concrete steps to be carried out, and not be too large.

Strengths of Heuristic Evaluation

Here are some strengths of heuristic evaluation:

- Its ease of implementation and high efficiency (Law et al., 2002; Nielsen, 1994b). It is considered to have a good success rate in that typically only three to five usability experts are needed to detect most (75–80%) of the usability flaws a system presents (Nielsen, 1994b).
- Its early applicability in the development lifecycle and low cost: it requires neither a working prototype nor the real users (Nielsen, 1994b).
- It is becoming part of the standard HCI curriculum and therefore known to many HCI practitioners (Greenberg et al., 1999). The heuristics are well documented and therefore easy to learn and put to use, so it may be argued that heuristic evaluation can also be effectively conducted by non-usability experts (Nielsen, 1994b).

On the whole, heuristic evaluation is considered to be a cost-effective evaluation method. Its main strengths lie in providing discovery and analysis resources (Cockton et al., 2003), such as domain and system knowledge, where it generally outperforms other popular inspection techniques like guideline-based methods or cognitive walkthrough (Wharton et al., 1994).

Limitations of Heuristic Evaluation

Here are some specific limitations of heuristic evaluation:

- Heuristic evaluation is highly dependent on the skills and experience of the specific usability expert(s) involved. At a high level of generality, the heuristics are "motherhood statements that serve only to guide the inspection rather than prescribe it" (Greenberg et al., 1999).
- Participants are not the real users. Regardless of the experts' skills and experience, they are still "surrogate users" (i.e., experts who emulate real users) (Kantner et al., 1997), therefore the resulting data are not really representative of the real users.
- Heuristic evaluation does not fully capture or take into account the context of use of the system under evaluation but rather evaluates it "as a relatively self-contained object" (Muller et al., 1995).
- It has been said that the majority of usability flaws detected by heuristic evaluation are 'minor' usability problems (for instance, by Nielsen, 1994a), or false positives, problems that do not negatively impact user performance or users' perception of system quality (Simeral & Russell, 1997).

When compared to other expert techniques such as guideline-based methods and cognitive walkthrough, heuristic evaluation is strong in terms of thoroughness (percentage of problems found), but weak in terms of efficiency (number of true positives vs. false positives) and, like other inspection methods, is vulnerable to expert biases (Cockton et al., 2003).

APPROPRIATING USABILITY HEURISTICS

Analyses of HE have shown that it is more likely for this method to miss relevant usability problems when the system to be evaluated is highly domain-dependent, and when evaluators have little domain expertise[2]. To overcome these limitations of the method when applied to mobile systems and settings, we have conducted an in-depth investigation of usability issues affecting mobile applications. The work leading to the set of specialized heuristics for mobile computing presented in Section 4.2 is based on this empirical evidence. The goal of the mobile heuristics described in Section 4.2 is to better support and contribute to the domain expertise of evaluators applying HE to mobile computing.

Methodology for Realizing Mobile Issues

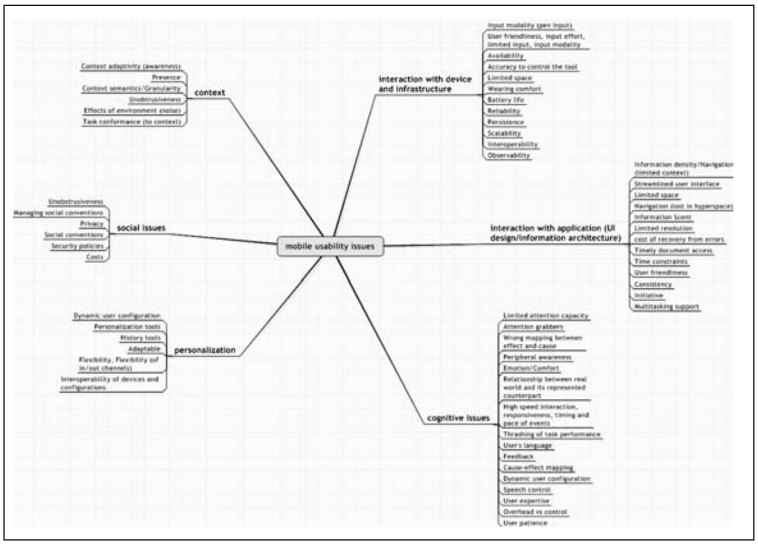
To develop usability heuristics for mobile computing, three authors of this chapter worked as usability researchers at the following activities:

1. Each of the three was assigned a unique set of papers to analyze independently. The papers originated from the list used in Kjeldskov et al. (2003); a recent meta-analysis of HCI research methods in mobile HCI[3]. The list was updated with papers published in the period 2004–2005 and selected only those with elements of evaluation. The analysis entailed documenting; for each of the papers, appropriate values for the following dimensions:

   - **Evaluation goal:** The evaluation mainly intended to demonstrate whether one technique is better than another, that is, a comparative study; or is it mainly exploratory, that is, understanding what kind of usability problems may rise with a given design.
   - **Evaluation method:** The evaluation method expert-based (made by experts through inspection), user-based (observing users performing tasks), or model-based (computing usability metrics through formal models).
   - **Evaluation setting:** The evaluation conducted in a laboratory (or any other controlled setting) or in the field.
   - **Real device/emulator:** The application under inspection tested with a real device or in a emulated environment.
   - **Location matters:** Does the application take location into account or not.

   Moreover, each of the usability researchers individually documented mobile usability issues that were indicated by (or evident from) each of the papers. At the end of this process, three different lists of usability issues were produced, containing the analysis of all the papers collected in the first phase.
2. In the next step, the usability researchers came together and consolidated their individual realizations. Individual findings had to be cross-checked and merged into a single consolidated list. This was done in the form of a spreadsheet.
3. Each researcher was then given the same realized list of mobile usability issues and asked to independently categorize (group or cluster) the issues. The idea was to find a way to summarize all the encountered issues and present them at a higher level of abstraction. On reflection, this was useful to check whether traditional heuristics covered each class of usability problems, or not. The researchers then presented and shared their individual categorization results with each other. Each researcher was requested to individually work further on his/her categorization with reference to the other categorizations, by eliminating redundant usability issues, clarifying the mobile usability issues, and grouping the obtained issues to an abstraction level that would be appropriate for developing/generating heuristics. Finally, they came together again to brainstorm and consolidate their work,

Figure 1. Mobile usability issues



and to harmonize the terminology used to describe the issues. Figure 1 shows an image of the set of issues produced. Here the top-level classes of problems are presented with examples of subclasses to make their meaning clear.

Mobile Usability Issues

The usability issues collected and grouped in high-level classes as described in the section on methodology for realizing mobile issues is described.

- **Interaction with device and infrastructure:** Many of the problems found in the research have a strong connection with the limits of the device and/or the infrastructure it is connected to.
- **Interaction with application:** This collects classes of problems connected to traditional screen design and information architecture.
- **Cognitive issues:** Here usability problems stemming from an overload of cognitive resources or a mismatch between a cognitive model and reality are characterized. While this aspect has always been taken into account in traditional studies, in mobile settings it becomes more evident and presents new challenges.
- **Personalization:** Standard heuristics tend to overlook problems connected to personalization or adaptation. While in standard settings this issue can be considered minor with respect to others, with mobile devices this aspect can really be critical.
- **Social issues:** Mobile devices and applications are used in a wide spectrum of environments and social conditions: private or public, alone or in groups, and so forth. This means that the social impact of adopted design solutions cannot be underestimated. Issues like privacy, security, user image, and social conventions thus become of great importance.
- **Context:** Similarly to social issues, it is necessary to take into account how the environment can affect interaction. Not only do social conventions and relationships with people matter, but also how potential physical environment features affect the design of an interface.

Methodology for Realizing Mobile Heuristics

This section discusses the research toward developing a set of mobile usability heuristics and also the efforts toward assessing the proposed mobile usability heuristics.

Toward a Set of Heuristics

The brainstorming activity described in the section on methodology for realizing mobile issues was continued and further articulated in a series of new individual or collaborative tasks aimed at developing a set of heuristics for mobile computing evaluations. By capitalizing on the outcome of the previous analysis of mobile usability issues, it was decided to rely on the following developmental process to come up with a new set of heuristics, better suited to be applied to mobile evaluation settings.

Phase 1

Each of the three usability researchers was provided with a table reporting Nielsen's traditional heuristics (Nielsen, 1994b) together with their corresponding definitions. Each researcher worked individually at assessing: which of Nielsen's heuristics were considered irrelevant for mobile settings; which of Nielsen's heuristics were relevant, but needed some revision or modification; and which additional heuristics needed to be included in the original set to cover relevant aspects of mobile applications. To better steer the individual relevance judgment of the heuristics, it was thought to be useful to define a guiding principle to be adopted and shared during the assessment work: this was a concise answer to the question: "What are the primary goals of mobile applications?," which was expressed as follows: "To enable a user-friendly navigation of relevant information or features in mobile conditions of use." The assessment and brainstorming activity performed in this phase was also informed by the consolidated version of the mobile usability issues that had been previously realized (the section on methodology for realizing mobile users).

Phase 2

Each of the usability researchers compared her/his own table of proposed heuristics with that of another researcher, to produce a new consolidated table. This activity was meant to be carried out individually, but based on comparing the work done by two researchers. The aim was to speed up the improvement of the set of heuristics proposed, in terms of their clarity and relevance to the mobile application field.

Phase 3

A new refinement process was started on the set of heuristics included in the three consolidated tables produced in phase 2. It involved: first, a discussion meeting among the usability researchers to arrive at a shared table consolidated from the three developed in phase 2; then, submitting this set of heuristics (with their definitions) to a number of targeted HCI researchers and professionals in the mobile computing and usability

Table 1. Mobile usability heuristics

Mobile Heuristic	Description
Heuristic 1	Visibility of system status and losability/findability of the mobile device
Heuristic 2	Match between system and the real world
Heuristic 3	Consistency and mapping
Heuristic 4	Good ergonomics and minimalist design
Heuristic 5	Ease of input, screen readability and glancability
Heuristic 6	Flexibility, efficiency of use and personalization
Heuristic 7	Aesthetic, privacy and social conventions
Heuristic 8	Realistic error management

community, to elicit feedback on the adequacy of the heuristics proposed. Nineteen experts were contacted in person, by e-mail, or by phone, and feedback from eight of them was received. The three researchers then met to discuss and compare the experts' comments with the researchers' consolidated table and arrived at the final set of mobile usability heuristics summarized in Table 1 and described:

- **Heuristic 1—Visibility of system status and losability/findability of the mobile device:** Through the mobile device, the system should always keep users informed about what is going on. Moreover, the system should prioritize messages regarding critical and contextual information such as battery status, network status, environmental conditions, and so forth. Since mobile devices often get lost, adequate measures such as encryption of data should be taken to minimize loss. If the device is misplaced, the device, system, or application should make it easy to recover it.
- **Heuristic 2—Match between system and the real world:** Enable the mobile user to interpret the information provided correctly, by making it appear in a natural and logical order; whenever possible, the system should have the capability to sense its environment and adapt the presentation of information accordingly.
- **Heuristic 3—Consistency and mapping:** The user's conceptual model of the possible function/interaction with the mobile device or system should be consistent with the context. It is especially crucial that there be a consistent mapping between user actions/interactions (on the device buttons and controls) and the corresponding real tasks (e.g., navigation in the real world).
- **Heuristic 4—Good ergonomics and minimalist design:** Mobile devices should be easy and comfortable to hold/carry along as well as robust to damage (from environmental agents). Also, since screen real estate is a scarce resource, use it with parsimony. Dialogues should not contain information that is irrelevant or rarely needed.
- **Heuristic 5—Ease of input, screen readability, and glaceability:** Mobile systems should provide easy ways to input data, possibly reducing or avoiding the need for the user to use both hands. Screen content should be easy to read and navigate through notwithstanding different light conditions. Ideally, the mobile user should be able to quickly get the crucial information from the system by glancing at it.
- **Heuristic 6—Flexibility, efficiency of use, and personalization:** Allow mobile users to tailor/personalize frequent actions, as well as to dynamically configure the system according to contextual needs. Whenever possible, the system should support and suggest system-based customization if such would be crucial or beneficial.
- **Heuristic 7—Aesthetic, privacy, and social conventions:** Take aesthetic and emotional aspects of the mobile device and system use into account. Make sure that users' data is kept private and safe. Mobile interaction with the system should be comfortable and respectful of social conventions.
- **Heuristic 8—Realistic error management:** Shield mobile users from errors. When an error occurs, help users to recognize, to diagnose, if possible, to recover from the error. Mobile computing error messages should be plain and precise. Constructively suggest a solution (which could also include hints, appropriate FAQs, etc.). If there is no solution to the error or if the error would have negligible effect, enable the user to gracefully cope with the error.

Assessing Heuristics Performance

To investigate the potential benefits of applying the set of heuristics for the evaluation of mobile applications, an experimental study aimed at comparing the support provided by the new set of mobile heuristics vs. standard usability heuristics (here, Nielsen's heuristics) to experts performing heuristic evaluation of mobile applications was devised and conducted.

Experimental Design

Here, various parameters pertaining to the set-up or design of the experimental study is described.

Participants and Materials

The study enlisted eight usability experts[4], as participants, to perform a heuristic evaluation (HE) of two mobile applications for which a number of usability flaws had already been identified. The two criteria used to select the applications to test were: being a typical application whose problems are known and evident; application whose tasks are simple and/or self-evident. After searching for applications fulfilling the foregoing conditions, the following two applications were chosen: Appl.1) a mobile device application in which location matters or that primarily relies on mobility: a PDA-based supermarket application was considered; Appl.2) a mobile device application in which interface navigation is key: A Web-based freeware e-mail application for PDAs[5] was considered. The following materials for the evaluators was also prepared: consent form, demographics questionnaire, post-evaluation form for participant's comments (to be filled out by the study moderator), a set of Nielsen's 10 usability heuristics, the proposed set of mobile usability heuristics (Table 1), and Nielsen's five-point Severity Ranking Scale (SRS) (Nielsen, 1994b) (which is described in Table 2).

Experimental Conditions

The experiment had the following two experimental conditions:

- **Condition 1:** N. 4 experts individually performed the HE by applying Nielsen's standard set of heuristics and Nielsen's SRS to both applications.
- **Condition 2:** N. 4 experts individually performed the HE by applying our set of mobile heuristics and Nielsen's SRS to both applications.

Procedure

The eight usability experts were randomly split into two groups, each assigned to one of the foregoing two experimental conditions (that is a between-subjects design). They all had previous expertise in the HCI evaluation field and were familiar with both the application of traditional HE methods and the use of mobile applications. Nevertheless, they were all given some brief instruction on the technique before starting the evaluation. The following protocol was used for both experimental conditions:

- **Pre-evaluation session:** This entailed first welcoming and greeting each evaluator. After that the goals of the study, the testing procedures, and the confidentiality issues were explained in detail. Scripts were prepared in advance and used for each usability evaluator to ensure consistency across

Table 2. Severity ranking scale (SRS)

Rating	*Description*
0	I don't agree that this is a usability problem at all
1	Cosmetic problem only. Need not be fixed unless extra time is available on project
2	Minor usability problem. Fixing this should be given low priority
3	Major usability problem. Important to fix, so should be given high Priority
4	Usability catastrophes. Imperative to fix this before product can be released

experts and conditions. In a demographics questionnaire experts were asked about their level of education/academic status, relevant experience in both HCI and mobile computing, experience in using both a PDA and Nielsen's heuristic evaluation method; the collected demographic data can be seen in Table 3. Most of the participants have a high level of education and an average knowledge of HCI and mobile devices. Six participants consider themselves almost knowledgeable about heuristic evaluation, while two give themselves an average rating. A training session was conducted with each evaluator to ensure that they fully understood the usability heuristics, and especially the mobile heuristics, which the participants were not familiar with; this involved the facilitator stepping through each usability heuristic and inviting the evaluators to ask questions in order to clarify the meaning of each heuristic and their understanding of the overall process.

- **Evaluation session:** The usability evaluators performed the usability evaluation on the mobile device by identifying usability problems and prioritizing them according to Nielsen's SRS (Table 2). Presentation of the two applications to be evaluated was counterbalanced to avoid any order effect. While evaluating the mobile device, each usability evaluator was asked to 'think aloud' to explain what s/he was trying to do and to describe why s/he was taking the action. Their comments were recorded by one of the evaluation moderators.
- **Debriefing session:** This focused on the evaluators' experiences of the process, and provided an opportunity to probe where behavior was implicit or puzzling to the researchers.

Table 3. Participants' demographics. Each value is ranked on a scale between 1 (min) and 4 (max)

Part.	Edu	HCI	PDAs	HE
p1	1	2	3	3
p2	4	2	2	3
p3	4	3	3	3
p4	1	2	2	3
p5	3	3	3	3
p6	2	2	1	2
p7	3	3	2	3
p8	1	1	1	2

Data Analysis

The data collected were analyzed both qualitatively and quantitatively. Comparison of HE effectiveness in the two experimental conditions was assessed.

Number of Flaws and Variation Among Experts

From Table 4, it appears that the use of the mobile heuristics has increased the number of flaws identified in the analysis of both applications, and has reduced variation among experts' analyses. In comparing the type of flaws detected by using the two different sets of heuristics, evidence of problems identified only by using Nielsen's heuristics was not found. The additional flaws found by applying mobile heuristics were usually different from the ones identified by using Nielsen's heuristics; also, the problems identified by each expert in the mobile heuristics condition were a small number from a larger set of usability difficulties presented by the two applications, although some overlaps were found (problems pointed out by more than one expert), which supports the idea of inter-expert consistency when applying mobile heuristics.

Severity of Flaws and Distribution

As depicted by Table 5 and Figure 2, Nielsen's heuristics have produced a more equally distributed severity ranking of problems detected for both applications. On the other hand, the mobile heuristics have produced a more positive evaluation of Appl.1 (61% of problems are considered minor or cosmetic) while for Appl.2 the ranking

Table 4. Number of usability problems identified

	Appl. 1	Appl. 2	Total	Mean (SD)
NHE	22	28	**50**	**12.5 (10.40)**
MHE	26	38	**64**	**16 (3.74)**

Table 5. Actual number of flaws and severity

	NHE			MHE		
	Appl. 1	**Appl. 2**	**Mean % for both appl.s**	**Appl. 1**	**Appl. 2**	**Mean % for both appl.s**
Cosmetic	5	7	6	5	5.75	5.375
Minor	5.5	6.5	6	11	13.25	12.125
Major	6.5	6.5	6.5	9	10.25	9.625
Catastrophe	5	8	6.5	1	8.75	4.875
Total of flaws	**22**	**28**		**26**	**38**	

seems to be equally distributed among the four severity levels. Considering the mean values in Figure 2, it does appear that Nielsen's heuristics identify fewer minor and major flaws compared to the mobile heuristics. It also seems that Nielsen's heuristics have a relatively even distribution of severity ranking for the problems identified. Nielsen's heuristics could therefore do a moderate job of identifying flaws at any design level. The mobile heuristics do seem to be especially good at identifying minor and major flaws rather than those at the extremes.

Figure 3 can be used for further analysis of how the specific heuristics from both sets fare with regard to average severity and average number of flaws. Figure 3 indicates that mobile heuristics are more effective in supporting the detection of flaws, while Nielsen's heuristics seem better suited to cover the case in which high severity flaws are present; also, mobile heuristics seem to support a more detailed evaluation of the mobile application (without considering the flaws classified as catastrophic). It is worth noting that some of the foregoing observations from Figure 3 are similar to those from Figure 2.

Figure 2. Actual number of flaws, severity and distribution

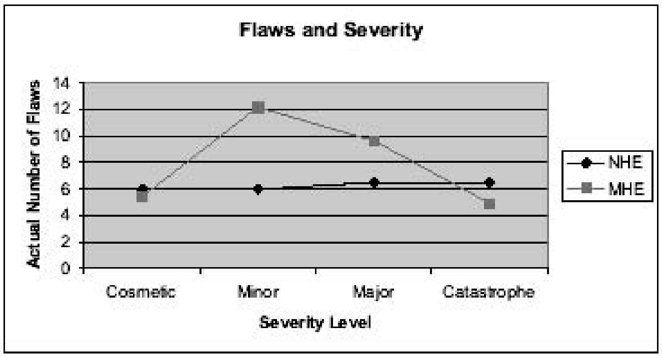
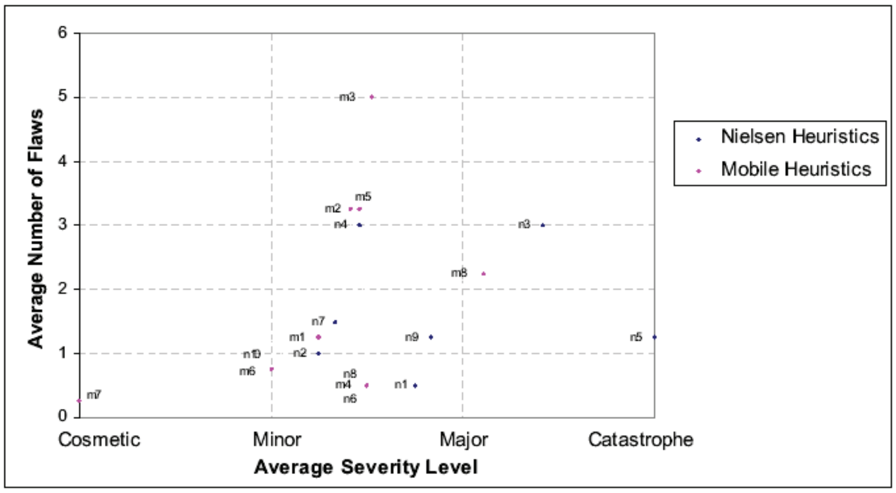


Figure 3. Comparison of the sets of heuristics: flaws and severity



So far it might be observed that the mobile heuristics produce a more accurate evaluation in terms of the number of problems detected (more flaws are identified), reduced variation among experts' analyses, and problems' severity ranking (this is actually also supported by the qualitative data collected during the evaluation, where most experts said that Appl.1 was much better designed for a mobile use when compared to Appl.2). Thus the mobile heuristics tend to focus the evaluation on the mobile issues instead of directing experts' attention at a more general level (although the kind of setting used in this study promoted an evaluation of applications that was more functionalities-based than contextual). Moreover, the mobile heuristics could be applied when/where the extreme flaws have been addressed or are not an issue in the design. If such flaws have to be identified before proceeding, mobile heuristics could be applied after Nielsen's heuristics. It is worth recalling that there are some problems that Nielsen's heuristics failed to identify (based on Table 4). Some might now be identified by mobile heuristics and might lie between minor and major severity levels (Table 5 and Figure 2).

Usability Flaws and Heuristics

As seen in Table 6 and Table 7, the most frequently used/highlighted heuristics in the mobile applications are as follows[6]:

- **Nielsen's heuristics:** Nielsen's heuristic 4 (12 times), Nielsen's heuristic 3 (12 times). The foregoing are [each] less than any of the following mobile heuristics.
- **Mobile heuristics:** Mobile heuristic 3 (20 times), mobile heuristic 5 (13 times), mobile heuristic 2 (13 times).

It is interesting to observe that these highlighted Nielsen's heuristics (4 [Consistency and standards], 3 [User control and freedom]) are related to the highlighted mobile heuristics (3 [Consistency and mapping], 5 [Ease of input, screen readability, and glaceability], 2 [Match between system and the real world]). The highlighted Nielsen's and mobile heuristics could be considered as being those most violated or most noticed, although their recurrence could be due to the particular type of evaluation/application(s) that were provided to experts.

The mobile heuristics probably scored such high figures (i.e., were able to identify more flaws under these related heuristics) because of the way

Table 6. Nielsen's heuristics and corresponding usability problems

Nielsen's Heuristics 1-10	Number of Usability Problems	Description of Heuristic
4	12	Consistency and standards
3	12	User control and freedom
7	6	Flexibility and efficiency of use
9	5	Help users recognize, diagnose, and recover from errors
5	5	Error prevention
2	4	Match between system and the real world
10	3	Help and documentation
1	2	Visibility of system status
6	2	Recognition rather than recall
8	2	Aesthetic and minimalist design

Table 7. Mobile heuristics and corresponding usability problems

Mobile Heuristics 1-8	Number of Usability Problems	Description of Heuristic
3	20	Consistency and mapping
5	13	Ease of input, screen readability, and glaceability
2	13	Match between system and the real world
8	9	Realistic error management
1	5	Visibility of system status and losability/findability of the mobile device
6	3	Flexibility, efficiency of use, and personalization
4	2	Good ergonomics and minimalist design
7	1	Aesthetic, privacy, and social conventions

the mobile heuristics have been revised and/or extended to capture mobile computing aspects.

It is interesting to note that the mobile heuristics "bring to the top" heuristics that are related to context. For instance: Nielsen's heuristic 2 has a score of 4; the related revised heuristic for mobile computing (mobile heuristic 2) scores 13. It may therefore be observed that the mobile heuristics make issues and flaws that have to do with context more apparent during the evaluation. Also, from the qualitative analysis of experts' reports, it was found that when the evaluator identified a flaw that could not be straightforwardly mapped to a specific mobile heuristic, s/he chose to assign it to mobile heuristic 2 or 3. Moreover, an evaluator stressed the need to make more explicit the word 'context' in the description of mobile heuristic 3. The description of the heuristic is found in the section on methodology for realizing mobile heuristics.

As seen in Table 6, the participants reported some usability problems regarding 'Help and documentation' (Nielsen's heuristic 10). This observation may be an indication that people using mobile applications still expect such applications to provide help. Though they might prefer that the help be 'interactive', non-distractive, not be a separate task, and so forth, the designer could consider the use of audio or some 'light-weight' approach (e.g., FAQs).

Time Taken to Evaluate

It seems that the application of the mobile heuristics was more time demanding during the whole evaluation, as seen in Table 8. This may be due to the experts' relative unfamiliarity with these heuristics compared to Nielsen's heuristics. This familiarity issue was tried to be reduced (it cannot be eliminated simply in an evaluation session) by giving the experts who were using the mobile heuristics some extra time at the beginning of the evaluation to study the mobile heuristics, in order to familiarize themselves with them and to ask questions. Although the application of the mobile heuristics was more time demanding, it should, however, also be observed that variation among experts was relatively high, confirming that heuristic evaluation is an evaluation technique strongly dependent on experts' previous knowledge and expertise with the heuristics, the application domain, and so forth.

As a general observation, it is worth mentioning that because the study adopted a between-subjects design, with the inherent risk that individual differences between participants can bias results, the fact that the application of mobile heuristics results in reduced variation among the participants' analyses is therefore notable.

REFLECTIONS TOWARD DEEPER PRINCIPLES

Many of the heuristics in 'traditional' Heuristic Evaluation appear to be phrased in a way that is general over all systems, and this is true also of many of the other forms of design guidelines or principles used in interaction design. This raises a number of questions:

Table 8. Time taken in minutes

	Appl. 1	Appl. 2	Total	Mean (SD)
NHE	106	92	**198**	**49.5 (27.196)**
MHE	155	136	**291**	**72.75 (44.776)**

1. Why do we need specific heuristics for mobile devices—why not use standard ones?
   If (as this chapter assumes) we do need them, then this raises further questions:
2. How do we know the heuristics that have been presented are correct?
3. How do we know they are sufficient or complete?
4. How do we know they work in real design?
5. Can we assess their scope—do the heuristics simply reflect current mobile technology and applications?

The answer to (1) is that we do need specific heuristics because the traditional heuristics implicitly embody assumptions about static desktop location and use. The differences between standard heuristics and the mobile usability heuristics presented in the section on toward a set of heuristics, are precisely due to the differences between mobile and fixed use.

The confidence on the correctness (2) and sufficiency (3) is based on the rigorous methodology used to derive the heuristics, distilling the knowledge and expertise in published work (the section on methodology for realizing mobile users), more analytic refinement of established heuristics (the section on toward a set of heuristics, phase 1), and review by experts (the section on toward a set of heuristics, phase 3). The feedback from experts gives some confidence in utility (4), and this was confirmed by the empirical study (the section on assessing heuristics performance), which also bolstered confidence in the correctness and sufficiency.

No set of heuristics or guidelines will be complete, but it can be sufficient to cover the more common or serious pitfalls. However, while the process of distillation from expert opinion and empirical testing suggests that the heuristics are sufficient for current mobile applications, on their own they do not tell us about applicability in the future (5). Mobile technology is changing rapidly and new applications are emerging. While it would be foolish to believe that all the ramifications of these can be forseen, it can be tried to ensure that the scope of the new mobile heuristics is understood. In particular, if it is attempted to make explicit the different assumptions that underlie the new heuristics, it will be able to be told when these change and will be in a better position to add to or modify this set in the future.

The differing assumptions that underlie desktop and mobile use under four headings will be examined: the nature of mobile devices, the environment of mobile infrastructure, the context of mobile use, and the purpose of mobile tasks. For each, the extent to which they are reflected in the heuristics and, where appropriate, how they may develop in the medium term will be examined.

The Nature of Mobile Devices

One of the most obvious differences between mobile devices and fixed ones is size.

It has been pointed out that desktop screen design is often lazy design—putting everything on screen and letting the user worry about what is important (Dix, 1999). In contrast, for the small screen of a mobile device, it is crucial that just the right information and input options are available at the right time—mobile device designers have to think far more carefully about the user's task than desktop designers. This is emphasized in Heuristic 4 "since screen real estate is a scarce resource, use it with parsimony. Dialogues should not contain information that is irrelevant or rarely needed." On a desktop application everything would just be shown (lazily). This is also reflected in Heuristic 2: "Enable the mobile user to interpret the information provided correctly, by making it appear in a natural and logical order," and in Heuristic 6: "Allow mobile users to tailor/personalize frequent actions…" While these are both good advice for any interface, it is particularly important on a small screen to help deliver the right information at the right time.

Several heuristics pick out issues of system adaptation. The system should (Heuristic 1) "prioritize messages," (Heuristic 2) "… sense its environment and adapt the presentation of information accordingly," (Heuristic 6) "suggest system-based customization," and (Heuristic 8), "constructively suggest a solution" for errors. In desktop systems, 'intelligent' system features can often get in the way and it is often better to have simple consistent interfaces. Of course this consistency is itself also more important when descriptions of actions are by their nature more parsimonious and Heuristic 3 focuses on this "consistent mapping between user actions/interactions." The balance between consistency and intelligence changes as the input/output bandwidth diminishes and the potential annoyance of wrong adaptations may be less problematic than the cost of doing everything by hand. Note too that Heuristic 3 is as much about *external* consistency with the environment as *internal* consistency over time.

The overall small physical size is also central to Heuristic 1's focus on losability/findability. A small device can easily get lost both in public places and in the home. However, its size means that it is often kept close at hand, both allowing it to be used as a proxy for the user in location services and also meaning that it becomes a very personal device, often used for private purposes. The importance of privacy is picked up in Heuristic 1: "Since mobile devices often get lost, adequate measures such as encryption of data should be taken to minimize loss" and Heuristic 7: "Make sure that users' data is kept private and safe." The personal nature is also picked up in Heuristic 7: "Take aesthetic and emotional aspects of the mobile device and system use into account." One of the unexpected lessons that mobile phone manufacturers had to learn quickly was that mobile phones are fashion items as well as functional devices. In addition, the content of mobile communications is often very rich and personal.

The Environment of Mobile Infrastructure

A key difference between driving across Africa and driving across Europe is the different transport infrastructure. The road system, signage, garages for repairs, and petrol filling stations are as much a part of the driving experience as the car dashboard. Similarly, for the mobile user the infrastructure in terms of wireless connectivity, charging, and data synchronization is as much a part of the mobile experience as the usability of the device itself.

The heuristics presented here are focused primarily on the use of the device itself and only marginally refer to this mobile infrastructure. Given that the heuristics reflect the current literature, clearly there is need for research in this area, which then may lead to further heuristics or guidelines for mobile infrastructure.

The influence of infrastructure can be thought of as different kinds of connectivity: connectivity to networks, connectivity to power, connectivity to data, and connectivity to location services.

Network connectivity is always of concern to mobile phone users and it is still not uncommon to see people hanging out of windows or waving phones in the air looking for signal. Heuristic 1 notes the importance of giving information on "network status" and phone users become proficient at reading the signal bars on their phones. In related technologies this is less well managed and owners of digital radio sets often become confused as digital stations seem to appear and disappear without warning; whereas analog broadcasts degrade slowly with distance, digital broadcasts can either be interpreted perfectly from weak signal, or not at all. Similarly, WiFi networks are seen as something akin to magic even by expert computer users, both in terms of how displayed signal levels correspond to actual access and in terms of the means to obtain connections through multiple levels of system property settings and authentication dialogues. Clearly, it will still be a long way from achieving even this simple goal of Heuristic 1.

The variability of network connectivity has been made a deliberate design feature in the notion of *seamful design* (Chalmers et al., 2004). Observations of network-based mobile games showed that players rapidly became aware that they could use patches of good or poor network connectivity in order to give them advantage during game play. This then led to games specifically designed using this notion.

Heuristic 8 on error management, while being partly general advice and partly about minimizing dialogue, is also indirectly related to connectivity. It is needed precisely because the user is far away from documentation, user guides, and expert help, and cannot rely on online help because of small screen size.

Heuristic 1 also notes the importance of "battery status." While battery technology has progressed remarkably, it is still one of the limiting factors for mobile devices, so much so that in some UK motorway service stations there are small racks of mobile phone lockers near the entrance where you can leave your phone to charge while you eat. The larger issues surrounding this are not mentioned in the heuristics, as a designer has little influence over them, but certainly standardization of power supply would seem an important step in reducing the plethora of power adaptors that so many of us carry while traveling, as well as making public power-charging facilities easier to manage.

Interestingly, power is not unrelated to network connectivity, as mobile phones consume more power if they have to connect to more distant radio masts. However, few users are aware of these interactions and an application of Heuristic 1 would be to give users a better feel for these things.

Heuristic 1 also notes the importance of minimizing data loss. This is related both to privacy (not losing data to others) and to data recovery. Data synchronization has a long history, back to early systems such as the CODA file system (Kistler et al., 1992; Braam, 1998), but still seems to be only poorly managed in practice. While there are ways to synchronize data between mobile devices and desktop systems, the fact that devices are connected through mobile networks could be used more widely to seamlessly backup crucial information such as phone address books. As mobile devices increase in data capacity, devices such as USB sticks and MP3 players are increasingly becoming the vectors for synchronization, so that perhaps new classes of application and hence a need for new heuristics will arise in this area.

Much of the early research on mobile platforms was based around more substantial mobile computers and (often collaborative) remote applications. Issues of data synchronization over networks were crucial and in particular, the problems due to network delays (Davies et al., 1996; Dix, 1995). This is still a major problem; for example, few users are aware that SMS is an unreliable messaging medium and can have substantial delays, especially for international texts. Strangely this does not seem to be prominent in the current literature and hence is not reflected in the heuristics. Perhaps this reflects the belief (as has been the case for 20 years) that 'soon' everything will be quick enough, or perhaps simply that empirical work is usually carried out in areas of high connectivity.

For location-based services, it is important that users understand the accuracy and other features of the location estimates. Where information or other services are based on discretized regions, confusion can arise at boundaries, rather like the digital radio example. Heuristic 1 applies again here, making not just location but data about uncertainty available. However, how this can be achieved in practice is still a matter for research, so more detailed general guidelines are not yet possible.

The Context of Mobile Use

Mobile devices are used while walking, (with care) in vehicles, outside in the rain, on the beach in the sand; they are often held while trying to do other things: open doors, carry shopping, pay for the bus, and in environments with other people. This rich set of physical and social contexts is reflected in several of the heuristics.

Heuristic 4 in particular, notes the importance of "good ergonomics" so that devices are "easy and comfortable to hold/carry along" and also that they are "robust to damage" when they inevitably get banged or dropped. Heuristic 1's focus on losability/findability reflects the dynamic context where mobile devices may be put down while carrying out other tasks.

The social context is also noted in Heuristic 7, "Mobile interaction with the system should be comfortable and respectful of social conventions" and this interacts with the ability to (Heuristic 6) "dynamically configure the system according to contextual needs."

Avoiding embarrassing symphonic ring tones and similar context-sensitive adaptations has been the focus of much research and a mobile device may potentially have access to just the environmental information to make this possible. While still very much a matter of ongoing research, this is reflected in several heuristics: (Heuristic 1) "messages regarding … environmental conditions," (Heuristic 2) "the system should have the capability to sense its environment," and (Heuristic 3) "user's conceptual model … consistent with the context."

Mobile devices are often used in far from optimal lighting conditions and while moving, making small fonts hard to read and so further reducing effective screen size. Both of these exacerbate the input/output problems of a small device discussed in the section on the nature of mobile devices. In addition, the user is often performing another task at the same time as using the mobile device, sometimes related to the mobile device's function (following a GPS map, or talking about a task on the phone) and sometimes unrelated (walking round the supermarket while talking to your mother on your mobile, or texting under the desk in a lecture). Heuristic 5 particularly picks up on these issues. When doing another task, it is essential that the user can "get the crucial information from the system by glancing at (*the device*)" and avoid "the need for the user to use both hands."

The Purpose of Mobile Tasks

Mobile applications can be split into two broad categories:

1. Those where location matters
2. Those where it does not

Many of the heuristics apply to both categories of application, but some apply more to one.

The first category includes location-aware applications such as navigation, tourist information, targeted advertising, and even augmented reality. Heuristic 2 talks about the "match between system and the real world" and says the system should "sense its environment and adapt the presentation of information accordingly." Heuristic 3 mentions the importance of consistency between users' actions and "the corresponding real tasks (e.g., navigation in the real world)." The ability of applications to achieve these aims is often dependent on hardware and environment so additional strategies are used, for example, the provision of several different landmarks to choose from, given an imprecise GPS location (e.g., Cheverst et al., 2000). Few location-aware systems include electronic compasses or gyroscopes, so directional consistency is particularly hard to maintain.

The second category is not simply the negation of the first, but the opposite: those applications that you specifically want to be able to do anywhere, such as being able to phone, access e-mail, read electronic documents, write in a word processor. In these applications the aim is to unshackle the user from the need to physically be in a particular place. Some of these are where the user wants access to remote resources from anywhere, and in this case issues of data and network connectivity may be important. For other applications in this category the data is local, but issues of screen size, portability, and dangers of data loss become more significant.

CONCLUSION AND FUTURE WORK

In this chapter, the benefit of expert-based evaluation methods and their need to capture contextual requirements in mobile computing have been pointed out. In the process, it has been described how mobile usability issues have been analyzed, and the efforts toward realizing a set of usability heuristics that is relevant to mobile computing have been discussed. The study confirms previous observations that mobile heuristics detect fewer cosmetic problems and that, in any case, they should not be considered as alternatives to user studies, but synergic. In particular, as often noted when speaking of inspection methods, it is believed that these are useful techniques to use when in early phases of design/prototyping, or when the low-cost issue is particularly relevant to the evaluation. As far as the false positives problem is concerned, the inter-expert consistency found when applying mobile heuristics may indicate that the flaws detected were not false alarms, although empirical evaluations with end users are the methods to uncover and solve this issue. As part of future work, it is intended to perform further literature analysis to the work reported in the section on methodology for realizing mobile issues and possibly consider more dimensions, and at different levels of abstraction.

REFERENCES

Braam, P. J. (1998). The Coda distributed file system. *Linux Journal, 50.*

Chalmers, M., & Galani, M. (2004). Seamful interweaving: Heterogenity in the theory and design of interac-tive systems. In *Proceedings of the ACM Designing Interactive Systems DIS'04.*

Cheverst, K., Davies, N., Mitchell, K., & Friday, A. (2000). Experiences of developing and deploying a context-aware tourist guide: The GUIDE project. In *Proceedings of MOBICOM'2000* (pp. 20-31). ACM Press.

Cockton, G., Lavery, D., & Woolrych, A. (2003). Inspection-based evaluations. In J. A. Jacko & A. Sears (Eds.), *The HCI handbook* (pp. 1118-1138). Lawrence Erl. Assoc.

Davies, N., Blair, G. S., Cheverst, K., & Friday, A. (1996). Supporting collaborative applications in a heterogeneous mobile environment, Special Issue of Computer Communications on Mobile Computing, 1996.

Dix, A. J. (1995). Cooperation without (reliable) communication: Interfaces for mobile applications. *Distributed Systems Engineering, 2*(3), 171-181.

Dix, A. J. (1999, December 15). Hazy, crazy, lazy days are over—time for designers to think (keynote). *Designing information for mobile and broadband network users*, London, Available at http://www.hiraeth.com/alan/papers/mobile-and-b-99/crazy-lazy-ohps.html

Greenberg, S., Fitzpatrick, G., Gutwin, C., & Kaplan, S. (1999). Adapting the locales framework for heuristic evaluation of groupware. In *Proceedings of the Conference of the Computer-Human Interaction Special Interest Group of the Human Factors and Ergonomics Society of Australia OZCHI'99.*

Johnson, P. (1998). Usability and mobility: Interaction on the move. In *Proceedings of theFirst Workshop on Human-Computer Interaction with Mobile Devices.*

Kantner, L., & Rosenbaum, S. (1997). Usability studies of www sites: Heuristic evaluation vs. laboratory testing. In *Proceedings of theInternational Conference on Computer Documentation SIGDOC '97* (pp. 153-160). ACM Press.

Karat, C-M., Campbell, R., & Fiegel, T. (1992). Comparison of empirical testing and walkthrough methods in user interface evaluation. In *Proceedings of the Conference on Human Factors in Computing Systems CHI'92* (pp. 397-404). ACM Press.

Kistler, J., & Satyanarayanan, M. (1992). Disconnected operation in the Coda file system. *ACM Transactions on Computer Systems.*

Kjeldskov, J., & Graham, C. (2003). A review of mobile HCI research methods. In L. Chittaro (Ed.), *International symposium on human computer interaction with mobile devices and services—Mobile HCI'03* (pp. 317-335). Springer-Verlag.

Kjeldskov, J., Graham, C., Pedell, S., Vetere, F., Howard, S., Balbo, S., et al. (2005). Evaluating the usability of a mobile guide: The influence of location, participants and resources. *Behaviour and Information Technology, 24*(1), 51-65.

Law, L. C., & Hvannberg, E. T. (2002). Complementarity and convergence of heuristic evaluation and usability test: A case study of universal brokerage platform. In *Proceedings of the Nordic Conference on Human-Computer Interaction NordiCHI '02* (pp. 71-80). ACM Press.

Molich, R., & Nielsen, J. (1990). Improving a human-computer dialogue. *Communications of the ACM, 33*(3), 338-348.

Muller, M. J., & McClard, A. (1995). Validating an extension to participatory heuristic evaluation: Quality of work and quality of work life. In *Proceedings of the Conference on Human Factors in Computing SystemsCHI'95* (pp. 115-116). ACM Press.

Nielsen, J. (1994a). Enhancing the explanatory power of usability heuristics. In *Proceedings of the Conference on Human Factors in Computing Systems CHI'94* (pp. 152-158). ACM Press.

Nielsen, J. (1994b). *Usability Engineering.* Morgan Kaufmann Publishers.

Nielsen, J., & Molich, R. (1990). Heuristic evaluation of user interfaces. In *Proceedings of theConference on Human Factors in Computing Systems CHI'90* (pp. 249-256). ACM Press.

Po, S. (2003). *Mobile usability testing and evaluation.* Master's thesis, University of Melbourne, Australia.

Sade, S., & Finland, O. N. (2002). Mobile ease-of-use and desirability through user-centered design. In *Proceedings of the SSGRR-2002 Summer Conference on Infrastructure for e-Business, e-Education, e-Science, and e-Medicine.*

Simeral, E. J., & Branaghan, R. J. (1997). A comparative analysis of heuristic and usability evaluation methods. In *STC Proceedings.*

Vetere, F., Howard, S., Pedell, S., & Balbo, S. (2003). Walking through mobile use: Novel heuristics and their application. In *Proceedings of the Conference of the Computer-Human Interaction Special Interest Group of the Human Factors and Ergonomics Society of Australia OZCHI2003.*

Wharton, C., Rieman, J., Lewis, C., & Polson, P. (1994). The cognitive walkthrough method: A practitioner's guide. In J. Nielsen & R. L. Mack (Eds.), *Usability Inspection methods* (pp. 105-140). New York: John Wiley & Sons.

KEY TERMS

Cognitive Walkthrough: Cognitive walkthrough is an expert-based evaluation method that is usually performed by experts in cognitive psychology. It evaluates the design on how well it supports the user in accomplishing a task.

Context: "Context is any information that can be used to characterize the situation of an entity. An entity is a person, place, or object that is considered relevant to the interaction between a user and an application, including the user and applications themselves." (Dey, A. Providing architectural support for building context-aware applications, Ph. D. Thesis Dissertation, College of Computing, Georgia Tech, December 2000). Context includes: location (absolute/relative, physical/virtual, etc.), infrastructure/resources (server and network capabilities and connections, applications), user (user data, usage patterns), environment (physical attributes such as light, temperature, humidity), entities (people, devices, objects), and time (date, time, season).

Expert-Based Evaluation: In expert-based evaluation, a designer or HCI expert assesses a design based on known/standard cognitive principles or empirical results. Expert-based evaluation techniques are also referred to as expert analysis techniques. Examples of expert analysis methods include: heuristic evaluation, cognitive walkthrough, and review-based evaluation.

Heuristic Evaluation: Heuristic evaluation is an expert-based evaluation method in which HCI or usability experts scrutinize the user interface and its elements against established design rules.

Mobile Device: A mobile device may mean one or more of the following: device that can be carried by the user (e.g., cell phone, PDA, wrist watch, etc), device that is autonomous (e.g., robot), or device (e.g., car computer) that is embedded in a moving object/device (e.g., car) (Dix, A., Rodden, T., Davies, N., Trevor, J., Friday, A., & Palfreyman, K. Exploiting space and location as a design framework for interactive mobile systems. ACM Transactions on Computer Human Interaction, 7(3), 285-321, September 2000). In this chapter, the first and the last are focused on.

Mobile/Ubiquitous Computing: Weiser coined the term ubiquitous computing. He also gave a vision of people and environments augmented with computational resources that provide information and services when and where desired (Weiser, M. The computer for the 21st century. Scientific American. 265(3), 91-104, 1991). Dix et al. define ubiquitous computing as: "Any computing activity that permits human interaction away from a single workstation." (Dix et al., 2003).

Review-Based Evaluation: Review-based evaluation is an expert-based evaluation method that relies on experimental results and empirical evidence from the literature (for instance from psychology, HCI, etc.) in order to support or refute parts of the user interface design.

Usability: Usability is: "The extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency, and satisfaction in a specified context of use." (ISO 9241-11 Ergonomic requirements for office work with visual display terminals: Guidance on usability, 1998). ISO 9241 identifies the following as the most useful indicators in measuring the level of usability of a product: effectiveness in use, efficiency in use, and satisfaction in use.

User-Based Evaluation: User-based evaluation is evaluation through user participation, that is, evaluation that involves the people for whom the system is intended: the users. User-based evaluation techniques include: experimental methods, observational methods, questionnaires, interviews, and physiological monitoring methods.

ENDNOTES

1 http://www.mais-project.it

2 See also http://www.useit.com/papers/heuristic/usability_problems.html

3 The papers were selected from top-level conferences and journals like CHI, AVI, UIST, TOCHI, and so forth. For details see J. Kjeldskov and C. Graham. A Review of Mobile HCI Research Methods. In L. Chittaro, editor, *International Symposium on Human Computer Interaction with Mobile Devices and Services—Mobile HCI'03*, pages 317–335. Springer-Verlag, 2003.

4 All the experts were new to the novel set of heuristics and none of the experts involved in the generation of heuristics discussed were involved in the experimental study.

5 HP iPAQ Pocket PC series h5500 PDAs with integrated wireless LAN (802.11b), 48 MB ROM, 128 MB RAM, and Intel processor 400 MHz were used. The PDAs were running Windows CE.

6 It should be noted that some of the participants indicated that some of the flaws were individually related to more than one type of heuristic (and thus the number of counts for the heuristics shown in Table 6 (and also Table 7) is greater than the number of flaws as shown in Table 4).

Chapter XLVII
Using Wizard of Oz to Evaluate Mobile Applications

Janet C. Read
University of Central Lancashire, UK

ABSTRACT

This chapter describes Wizard of Oz studies and gives a historical view that includes a summary of the literature in which several studies that used the method with mobile technology are discussed. The use of Wizard of Oz for mobile applications is explored by referencing the literature and by examination of a case study. A taxonomy for Wizard of Oz studies is presented that has been derived from a study of the literature in this area. A set of guidelines is presented that outlines the essential considerations in planning a Wizard of Oz study for mobile applications. The chapter concludes with some thoughts for future Wizard of Oz studies.

INTRODUCTION

In most technology applications, developers are required to evaluate systems for usability and user acceptance. These evaluations require careful planning and there are often pressures on resources and time that constrain the evaluation team. Where technology is particularly novel and where the cost of producing the technology is high, the evaluations that need to be carried out before any investment is provided pose a particular problem.

One method for evaluating novel or incomplete products is to use simulation. Simulation assumes that the evaluators engage with a product that 'looks like,' 'behaves like,' or 'feels like' the eventual product, but is, in one way or another, incomplete. The benefits of using simulations are that certain features of products can be evaluated before a fully functional product is available.

In mobile applications, evaluators are often concerned with the ease of interaction with a product as well as with the effect of the product on user behaviour. For many mobile applications, the development of the communication technology and the development of the product interface require considerable investment. However, for the purpose of evaluation of concept and evaluation of ease of use, these two components can be simulated and therefore tested early in the development stage keeping the cost of redesign low.

In HCI (Human Computer Interaction), evaluation using simulation has become synonymous with the term ‘Wizard of Oz’. A Wizard of Oz evaluation is one in which some, or all, of the interactivity that would normally be controlled by computer technology is ‘mimicked’ or ‘wizarded’. The Wizard of Oz method is considered mainstream in HCI and, as user groups have diversified and as the technologies under investigation have changed, the method has become more and more popular in evaluation studies.

Wizard of Oz methods have long been associated with novel and emerging interfaces. Originally presented as a method for simulating speech input at a time when speech recognition was flaky, the method is particularly well suited to mobile technology which is often novel and where, as indicated by Cohen and Oviatt (1994), ”portable computing and communications devices will soon be too small to allow for the use of a keyboard, implying that the input modalities for such machines will most likely be digitising pens and voice.”

This chapter presents an overview of Wizard of Oz as a method and then focuses down on its use for evaluations and specifically for evaluations of mobile applications. The first section introduces WOz (Wizard of Oz) methods by giving a historical overview which is augmented with references from early studies that used the method. This section concludes with a consideration of how WOz is used in the Product Development Lifecycle. The second section presents the literature on the use of WOz in mobile applications before concluding with a taxonomy that describes the variability of WOz studies. The third section focuses on the process of carrying out a WOz study, beginning with a case study that demonstrates how a WOz study was used in an evaluation of a mobile system in a museum and concluding with a discussion of when and how to use WOz. This section includes some tips and guidelines for the use of WOz with mobile technology. The chapter concludes with some reflections on the use of Wizard of Oz and presents some research ideas for future studies in this exciting area.

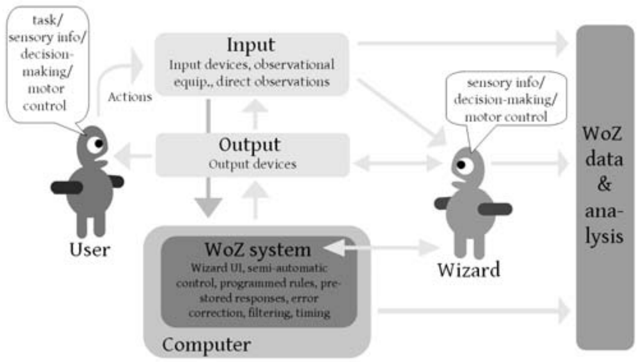
THE WIZARD OF OZ METHOD

A traditional Wizard of Oz study (shown in Figure 1) has three components; a human wizard, an interface, and a subject. During the study, the human wizard manipulates the interface in such a way that the subject is unaware (to varying extents) of the existence and the impact of the wizard.

There is some debate about the origins of the method as the concept was being used well ahead of the adoption of the name. As a method, it can be traced back to an IBM technical report by Thomas (1976), but it is more often attributed to Gould, Conti, and Hovanyecz (1983), whose study of a listening typewriter (an early simulation of a speech recognition system) is well cited and whose diagram of a WOz study is often reproduced. In the Gould et al. (1983) study, a skilled typist was employed to enter what the participants said to the computer and to therefore act as a wizard by mimicking the potential behaviour of a speech recognition engine.

The phrase Wizard of Oz comes from the book ‘*The Wonderful Wizard of Oz*’ by Baum (1900). Central to this book is the character known as the Wizard of Oz who is believed, by Dorothy Gale and her friends, to be the only one able to solve their problems. Unseen for most of the book, Oz is reluctant to meet the travellers and constantly appears in disguise, once as a giant head, once as a beautiful fairy, once as ball of fire, and once as a horrible monster. As the story progresses, it becomes apparent that Oz is actually none of these things, but is in fact just an ordinary American man who has been using a lot of elaborate magic tricks and props to make himself seem ‘great and powerful.’ Incidentally, his name, Oz comes from the first letters of the first two words in his name (Oscar Zoroaster).

Figure 1. Wizard of Oz setup (Höysniemi & Read, 2005)



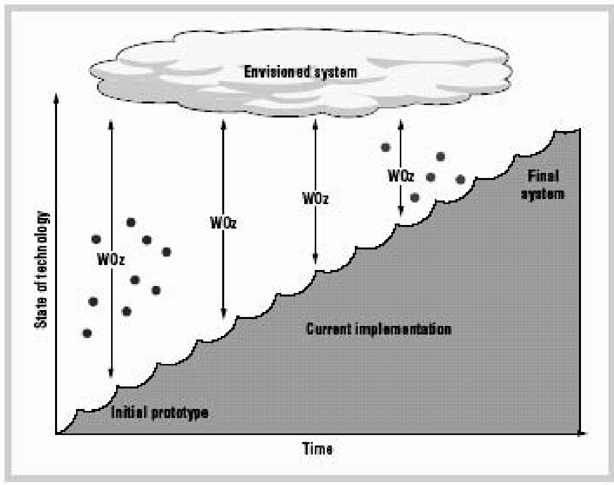
In work that described a natural language office information system, Kelley (1984) was probably the first to associate a simulation interface with the character Oz. He described his study as 'an experimental simulation which I call the OZ paradigm, in which experimental participants are given the impression that they are interacting with a program that understands English as well as another human would. In fact, at least in the earlier stages of development, the program is limping along, only partly implemented. The experimenter surreptitiously intercepts communications between participant and program, supplying answers and new inputs as needed.' It is not clear when the word Wizard was added to descriptions of Oz studies but, by the time Fraser & Gilbert (1991) had written one of the most comprehensive studies of the method, the phrase Wizard of Oz seemed prevalent.

Wizard of Oz in the Development Lifecycle

The Wizard of Oz method is extremely versatile as it can be used at several points throughout the system lifecycle (Dow et al., 2005). In the early stages, WOz can be used to gather requirements when all the developer has is a very sketchy idea of a system. Later on in the lifecycle, WOz studies can be used in the design of the system by gathering specific user needs in relation to a prototype system. This was the method employed in the WOz studies that gathered language corpora (Dahlback et al., 1993) and in more recent studies that have collected gestures and actions (Höysniemi, Hämäläinen, Turkki, & Rouvi, 2004; Paive et al., 2003). Once the system is specified, it is also possible to use WOz studies for evaluations, as was commonly the case in the early language studies (Gould et al., 1983).

When used for evaluation, WOz studies have generally been used to determine user perceptions of interfaces that, to all intents and purposes, present as finished systems to the user. These evaluations of user perceptions are especially important for those systems that that cannot easily be built (Gould et al., 1983; Oh et al., 2002). Another use for WOz evaluations is to control otherwise difficult to control situations. Examples of these are prevalent in the recognition field especially for studies that measure error tolerance in recognition based systems as described by Karam and Schraefel (2006), Read, MacFarlane, and Casey (2003), and LaLomia (1994). In these studies, whereas a fully functional system is possible, a wizard 'interrupts' the system and presents a 'programmed' response to users in order to observe the effect that this programmed response, rather than the uncontrolled response of the functional system, has on the user.

Figure 2. WOz in the lifecycle with permission from IEEE (Dow et al., 2005)



WIZARD OF OZ STUDIES FOR MOBILE TECHNOLOGY

Studies of Wizard of Oz for mobile systems fall into three categories. There are those that focus on the evaluation and design of the physical interfaces for mobile systems where the work is generally carried out in laboratory systems but the interface is intended to be used in a mobile application (for example the design of mobile phones and PDAs). There are studies that concern themselves with the design of interactions that could only be effected using mobile systems (where connectivity might be crucial) but these are also predominantly evaluated or designed in a lab based system for example, the calendaring system of Lyons, Skeels, and Starner (2005), and finally there are the location based systems where any wizard would have to be virtually or physically 'on the move' with the user as described in Li, Welbourne, and Landay (2006).

The first sort of study, interface design for mobile devices, generally concerns pen or speech based systems where the recognition aspect is the difficult feature that makes the study merit a Wizard of Oz approach. As such, these studies are not especially interesting as the 'mobility' of the interface is not a feature of the study.

One use of Wizard of Oz in an evaluation of a traditional mobile system (where connectivity affects the interaction) is described by Lyons et al.(2005). This study looked at speech that was intended to be used for human to human conversation but was also being used for spoken input to a computer system (referred to as dual purpose speech). Using PDAs, this study required users to make appointments during normal conversation. Thus, while a conversation took place, key words, like tomorrow, Friday or 12th, caused the system to navigate through a calendar application and enter appointments. To test the concept, a WOz study was devised that required the wizard to listen out for the key words and drive the interface. This study demonstrated the use of wizarding for applications that could be easily imagined but less easily built. Although intended for outdoor mobile use, this particular wizarded study was carried out indoors.

In another PDA study, Takayama, Leung, Jiang, and Hong (2003) used Wizard of Oz to evaluate three different interfaces for a PDA tracking system known as BuddySystem. In this instance, the wizard followed the users around a room whilst also holding a wireless laptop PC. The wizard updated location information to the PC which caused changes on the PDA interface as the user moved around.

Mobile applications based on location based systems are difficult to implement and, during testing, there are often problems with poor signals and wireless breakdowns. Both Dow et al. (2005) and Li et al. (2006) have used WOz studies to evaluate location based systems. In 'The voices of Oakland' study, Dow et al. (2005) used an augmented interface that supported both wizarded and non wizarded behaviour. This interface allowed the wizard to progress from controller to moderator to supporter as more and more functionality was added to the prototype system. As controller, the wizard watched the user and decided on all the actions, as moderator, the wizard could sort out, for instance, multiple button presses, and as supporter, he could observe the actions and overrule where necessary. The location based study reported by Li et al. (2006), adopted a different approach with the wizard following the user around the location under review. The wizard used an application on a tablet PC that allowed him to 'estimate' where the user was and provided some control over the user interface. In the discussion about the efficacy of the WOz method in this set up, Li et al. (2006) emphasised the difficulty for the wizard who had to track location, trigger application behaviours, and observe the environment all at the same time.

WOz studies have been used for some portable tangible systems which are sometimes embedded in mobile applications. In the SenToy project, Wizard of Oz was used to simulate a gesture recognition system that required users to manipulate a doll or teddy bear to convey emotions and actions (Paiva et al., 2002). In this study, users were told that the doll had sensors embedded in it and that by moving and manipulating it, the image on the screen would change. Whilst wizarding the interface, the experimenters were able to gather useful data about how the users manipulated the doll.

Variability in Wizard of Oz Studies

In reading about Wizard of Oz studies, it is apparent that there is a huge amount of variability in the WOz set ups. The process of carrying out a Wizard of Oz study can be referred to as 'Wozzing,' a term first used by Read, Mazzone, and Höysniemi (2005). Wozzing is a broad term that includes the design of the experiment, the design of the interfaces, the selection and training of wizards and participants, and the collection, analysis, and interpretation of the data. In the following list, some of the variability in Wozzing is highlighted and there is discussion of how this variability might impact Wizard of Oz studies with mobile applications.

- **The functionality of the technology:** Some studies use fully functional prototypes, others use low-tech or even non-tech prototypes. The functionality of the prototype is in part, determined by the stage in the lifecycle at which the WOz study is being done. In mobile applications, it is possible to have just the connectivity of the interface missing and for the user to see and experience what appears to be a complete system. Alternatively, the user could be talking to locations in a 'pseudo' augmented space that has absolutely no real functionality and no prototype.
- **The discretion of the wizard:** In some studies, the wizard is allowed to do whatever he pleases, in many studies the wizard is constrained to a set of options and occasionally, the wizard can be replaced with a robot (pre-programmed response) (termed a robot-of-oz study). If the desire is to find out what users do when faced, for instance, with a certain behaviour on a PDA, the wizard may be acting as a tight system and selecting from just a handful of options, or could be mimicking a natural language interface in which instance he or she could respond as suited.
- **The amount of wizard control:** The wizard may be the sole provider of functionality in the system or he may be only wizarding some of the interactions with the interface providing some automatic control. Clearly, if the interface does all the controlling, this would not be a Wizard of Oz method. Location based systems can be partially wizarded with the wizard providing actions based on real location information or the wizard track-

ing the location and allowing the computer to create the responses.

- **The visibility of the wizard:** The wizard may be seen or unseen. In a study by Read et al. (2005) it was shown that wizarding could easily be effected with the wizard in the same room as the users. One of the challenges with mobile wozzing is that in many cases, the interface is so small that the wizard needs to be very close to see what is going on.
- **The number of wizards:** In some studies there are multiple wizards carrying out different activities in a single study. In others, wizards are exchanged during the study (which will cause an effect, most obvious where the wizard has a high degree of autonomy). More commonly, there is one wizard carrying out all the action. For some studies it might be appropriate to have two wizards each doing different things; for example, in location based systems, one wizard could wizard the location and another wizard could deal with the system responses.
- **Wizard knowledge:** The wizard brings knowledge of the domain to the study. Some wizards are the system developers or designers; others are brought in for the experiment on an ad hoc basis and could be either experienced wizards or have very little experience. For instance, with a wizard simulating a speech input based system, if he or she has used real speech input based systems, the responses from that wizard may differ from those from a less knowledgeable wizard.
- **User knowledge:** Related to the visibility of the wizard is the amount that the user knows about the set up. Levels of deception vary across Wizard of Oz experiments from the user believing that all is being done by a functional interface to the user knowing that the wizard is doing all the manipulation. It is common to give the user knowledge that lies between these two extremes. According to how much the user knows about the set up, the study might be changed. A user that knows a wizard is running the process sometimes behaves differently.
- **Experimental design:** As with all experimental studies, Wizard of Oz studies vary from tightly controlled experiments to free exploration. With free exploration, the wizard either has to be highly trained or has to be mimicking a natural language interface and have a high degree of autonomy. It is very difficult to allow free exploration in most studies; one method is to have a wizard action that embraces variable and unknown commands. In many mobile applications, if the study is taking place outside, the design of the experiment may have to take account of interruptions and ad hoc interactions.

PLANNING A WIZARD OF OZ STUDY

As specified earlier, a Wizard of Oz study requires an interface (to a system), a wizard, and a user (often referred to in the literature as a subject). It is not possible to use Wizard of Oz studies for all systems. This section of the chapter discusses where it is appropriate to use WOz methods, what precautions need to be taken, and describes a case study that illustrates some of the points for consideration.

In some cases, the use of Wizard of Oz is limited by restrictions imposed by the system. Systems with non digital outputs (for example, pointing based drawing applications), systems that require rapid response (like some computer games or control systems), and systems that are difficult to observe (thought based computing for instance) are all problematic. Fraser and Gilbert (1991) specified that for a Wizard of Oz it must be possible to:

- Simulate the future system
- Specify the future system behaviour
- Make the simulation convincing

The first of these constraints is largely concerned with the technology. The other two constraints are much more concerned with the humans in the system. Humans in Wizard of Oz studies pose several limitations but these limitations have also provided the field with opportunities for innovative solutions.

The wizard, for instance, needs to be able to ‘mimic’ the behaviour of the system and therefore has to understand what the system would do. This requires that the system is not too complex and, for convincing simulation, the wizard needs to be left with not too much to do. Several ‘assistants’ for wizards have been developed to make the task easier and to reduce wizard error. For dialogue (question and answer) systems, where a wizard would have to remember computer responses, Dahlback, Jonsson, and Ahrenberg (1993) created ARNE-3, a simple menu based tool that assisted the wizard. Similar tools that have since been designed include SUEDE for speech prototyping (Sinha, Klemmer, Chen, Landay, & Chen, 2001), DiaWoz, a newer application for dialogue systems (Fiedler & Gabsdil, 2002) and OzLab for interface prototyping (Pettersson & Siponen, 2002).

As well as having assistive tools and instructions, the wizard needs to understand the context in which he or she is doing the wizarding and is likely to need some associated training. Gould et al.(1983) stressed the need for practise and expertness in their early study, reporting that the typist was both very skilled and very well practised. Training and expertness is also highlighted in Fraser and Gilbert (1991), who commented that the variability of the wizard with respect to response time, dialogue model (possibly reduced by an automated tool), and training was significant. In the same work, Fraser and Gilbert (1991) identified three areas in which the wizard should be trained; these being the application area, the system capability, and in the use of any assistive tools.

The user, or subject, in a wizard study is also a factor. Users are known to be variable and so the same precautions as would be taken in any study apply equally to a Wizard of Oz study but there are the additional dimensions of subject gullibility and knowledge of the set up. This user variability is not often well reported with studies generally constraining comments about the user to simply emphasise that the users did not mind the deception.

Case Study: Runaround

This study, a WOz evaluation by the author, was used in the design of a game that was to be played by children in a museum. The museum staff wanted a game that children would ‘play’ with physical objects in a room sized space which would help the children learn about the history associated with the objects. The first requirement, in the design of such a game, was to see if the concept was engaging.

The game concept that was agreed required children to move physical objects to hotspots at which point an audio snippet would play, offering some information. The audio would depend on the object and the hotspot. The room would be laid out like a world war two house and the objects would be world war two artefacts.

This game concept provided an ideal opportunity for a Wizard of Oz study: The outputs that would be expected were sounds which are very easy to manage in a wizarded environment, the interaction possibilities were limited, with children only expecting action if tagged item and hotspot came together, the space was easy to watch as there were plenty of hiding places and an additional person would create no great surprise, and the effort of programming an RFID tagged environment would be considerable. In addition, the ‘feature’ of the evaluation, engagement, was not expected to be affected by the lack of real interactivity.

Preparation for the WOz study began with a focus on the wizard. The cognitive load on the wizard was known to be related to the number of hotspots, the number of objects, and the amount of ambiguity. In addition, wizard performance would be affected by the rate at which new objects met with new hotspots, the need to learn interface actions, and time on task.

Pilot studies were used which determined that, given a medium sized room, and children aged between four and six, the wizard could manage six hotspots and eight objects. To simplify the wizard load, a special interface was constructed using VB6 that displayed the hotspots in an arrangement that could be easily changed to suit the view that the wizard had of the room, and, alongside each hotspot, showed images of the artefacts that the children were moving around. The wizard simply had to click the artefact in the appropriate hotspot location, to emit the correct audio file.

The evaluation took part in two stages. In stage one, the wizard practised the interface with four children recruited from a local school. In stage two, the formal evaluation, twelve children from a different class came singly to the museum space and were told that the tagged items and the hotspots created different things. They were told that the designers of the game were keen to 'see how good the game was.' The time the children spent on the game was recorded together with some engagement metrics.

During the evaluation there were several problems. The first was that the children sometimes had their backs to the wizard and so he had a poor view of what it was that the child had in his hand. This caused delays in the output which, although the children did not appear to have a problem with, might have had an impact on the resulting engagement scores. Another problem occurred where children played too quickly for the wizard, who lost his place on several occasions. This turned out to also be less of a problem than anticipated as, after the study, it transpired that the children were really only hearing the first few words of each sound output and were then moving on, to another physical action. A third problem occurred when one child brought all the objects to one hotspot; and the wizard was unsure what to do!

When the study was over, the children were brought back to the classroom and were told that the study had been wizarded. This provoked an unexpected behaviour when several of the children wanted to take turns at being the wizard with the result that all the children played the game again. Interestingly, with their peers doing the wizarding, the children took great delight in catching the wizard out!

Guidelines for WOz Studies for Mobile Applications

Returning to the literature from Fraser & Gilbert (1991), the developer or evaluator wanting to use a Wizard of Oz study with mobile technology needs to be confident that he or she can simulate the system, specify the system behaviour, and make the simulation convincing. If the system is fully functional, a WOz study is generally inappropriate but for all other studies WOz should be considered. The decision to use a WOz depends on what is required. Having decided to consider a WOz study, there are several stages to go through:

- **Decide what will be wizarded:** In deciding, consideration has to be given to the 'spy' method. For instance, if location tracking is being wizarded, methods will need to be devised to implement this. For simple 'mock ups' in the lab, the choice of features for wozzing will be easier.
- **Decide how many wizards will be used:** If the study is lengthy, wizards get tired and so there may be a need for more than one wizard. If there are many different things to watch, more than one wizard may be needed at the same time. If wizards are being swapped, ensure consistency by training and preparation.
- **Design a WOz interface for the user (if needed):** Sometimes the original interface can be used but in many instances, depending on the 'spy' methods, the interface used by the user may need to be modified.
- **Determine the 'spy' systems (the methods by which the wizard will know what the user is doing):** If cameras are being used, their position and their visibility need to be considered. If the wizard is being shown a capture of the screen (as is common in interface wizarding), the means to carry this from one room to another also needs to be resolved.
- **Determine the wizard behaviours:** For the intended system, the full set of user behaviours and user actions, and a corresponding set of wizard behaviours and actions needs to be mapped out. If there are only a handful of behaviours and the corresponding actions are simple, these can be presented in a tabular form for the wizard to refer to

during training with the assumption that he or she will easily learn them once practised. For more complex situations, either multiple user behaviours or complex wizard actions, the wizard will need assistance, either the recruitment of another wizard or the implementation of a wizard interface

- **Design a wizard interface (if needed):** Consideration should be given as to whether one of the ready made WOz interfaces could be used. If not, time has to be given to design a robust WOz interface and this will need user testing with the wizard. If there is visual information being transmitted to the wizard, this may need to be incorporated into, or presented separately from, the wizard interface.
- **Train the wizards:** As indicated by Gould et al. (1983), even expert wizards need training. The more complex the system, the more training the wizard will need. Aside from this training, the wizard may need instruction in the context and may need some information about the technical aspects of the simulated technology.
- **Pilot the study:** With a selection of users, pilot the study to ensure that all that needs to be captured is captured, that all the responses needed are in place and that the wizard is able to manage the study
- **Carry out the study:** Taking account of variable factors

Cautions for Wozzing

Wozzing is not without problems, these can be summarized as technology related problems, method problems, and wizard problems. These drawbacks are briefly expanded on in the next three sections.

Technology Related Problems

It is reasonable to assume that individuals that are deceived into thinking technology is able to do things that it cannot yet do, might become convinced otherwise. Unless the users are being brought back to do other related studies, they should be informed after the studies as to the real state of the simulated technology. If the technology is only being wozzed for convenience, as in the Runaround case study, the users can be shown an example of how the technology could be implemented.

Ethical Concerns with the Method

A major problem with WOz studies revolves around informed consent. Informed consent (Diener & Crandall, 1978) relies on individuals being well aware of what they are doing and in the case of WOz studies this is seldom the case. Indeed, informing the participants may spoil the WOz study. One way around this is to gather consent before and after the study, giving the participants the full information after the event and allowing them to withdraw their consent at that stage. If at all possible, open setups should be used where the wizard's role and actions are explained to the user before testing.

The Wizard is a Human

Many of the practical problems and concerns about WOz studies relate to the fact that the wizard is a human. The wizarding process puts the wizard under high cognitive and physical loading, observing the user's input, making decisions on what response is appropriate for the given input, and delivering system outputs using the available user interface. In addition to coping fluently with all 'normal' user input, the wizard has to be able to manage conflicting inputs, interpret boundary values, and adapt to the user's errors and their sometimes extraordinary behaviour.

There are several factors that influence the wizard's reaction time and decision-making. One of the key issues is the usability and natural controls of the WOz system. The number of input-output alternatives and their intuitive compatibility is crucial when the wizard has to act fast, especially in systems where a continuous interaction, for example in mobile applications, is required.

The wizard's motor abilities including reaction time, response orientation, speed of movement, and manual dexterity all affect the performance of the setup and this is a compelling reason for the wizard being well trained. Practising should result in faster reaction times, more correct anticipation of user's actions, and better recovery from errors. The heavy cognitive and physical load causes stress that leads to perceptual narrowing and reduces the attention span of the wizard. Because of this, test session lengths and breaks should be carefully designed into a WOz study. Where possible, the wizard should be recorded so that wizard errors and user errors can be separated.

CONCLUSION

There is clearly a lot of work still to be done to determine how useful WOz studies are for evaluations in all contexts and evaluations of mobile applications are no exception. There are very few studies that adequately report the use of WOz in mobile situations and the work on location based systems is very embryonic. Given the novelty of many mobile applications, and the high costs of development, it is likely that WOz will continue to be an important method for these systems. This may result in the development of more wizard interfaces for the effective management of these studies.

FUTURE TRENDS

There are several areas where further research is needed. One area is the effect of the wizards on the evaluation environment. Whilst it is generally recognised that the wizard has an effect, controlled studies that measure this effect are scarce. Variability across wizards and variability according to time on task are relatively unstudied.

As systems become more and more complex, the relationships between task complexity and wizard error are worth consideration. It may be possible to create a simple model that allows evaluators to predict wizard overload, making decisions about the number of wizards and the size of the evaluations easier to make.

In the consideration of ethical issues, studies with different user groups and in different contexts are needed to see how easy it is to carry out WOz with open systems where the user is aware of the wizard.

SUMMARY

As systems become more complex and the cost of implementation rises, there will be a greater need for cut-price methods for early investigation and evaluation and so it is expected that more, rather than less, WOz studies will take place. In addition, the use of WOz methods for requirements elicitation and for gathering user actions is likely to increase.

Wizard of Oz studies can appear deceptively simple. The promise of an evaluation without a system seems tantalisingly attractive but in reality, the work involved in designing good WOz studies is often greater than had there been a functional prototype. The reason for this is simple, in developing a WOz study, the designer has to predict and plan for all user actions and system responses.

Wizard of Oz methods are useful for the evaluation of mobile systems and, if well planned, can provide valuable information about user behaviour and the user experience that might otherwise be difficult to establish.

REFERENCES

Baum, L. F. (1900). *The wonderful Wizard of Oz.* Penguin Books Ltd.

Cohen, P. R., & Oviatt, S. L. (1994). The role of voice in human machine communication. In J. G. Wilpon & D. B. Roe. (Eds.), *Voice communication between humans and machines* (pp. 34 - 75). New York: National Academy pf Science.

Dahlback, N., Jonsson, A., & Ahrenberg, L. (1993). Wizard of Oz studies—Why and how. *Knowledge Based Systems, 6*(4), 258-266.

Diener, E., & Crandall, R. (1978). *Ethics in social and behavioural research*. Chicago: University of Chicago Press.

Dow, S., MacIntyre, B., Lee, J., Oezbek, C., Bolter, J. D., & Gandy, M. (2005). Wizard of Oz support throughout an iterative design process. *IEEE Pervasive Computing*, *4*(4), 18-26.

Fiedler, A., & Gabsdil, M. (2002). *Supporting progressive refinement of Wizard-of-Oz experiments*. Paper presented at the ITS 2002—Workshop on Empirical Methods for Tutorial Dialogue Systems, San Sebastian, Spain.

Fraser, N., & Gilbert, N. S. (1991). Simulating speech systems. *Computer Speech and Language*, *5*, 81-99.

Gould, J. D., Conti, J., & Hovanyecz, T. (1983). Composing letters with a simulated listening typewriter. *Communications of the ACM*, *26*(4), 295-308.

Höysniemi, J., Hämäläinen, P., Turkki, L., & Rouvi, T. (2004). Children's intuitive gestures in vision based action games. *Communications of the ACM*, 44-50.

Höysniemi, J., & Read, J. C. (2005). *Wizard of Oz with children—Ethical dilemmas and experiences from the field*. Paper presented at the IDC2005, Boulder, Colorado.

Karam, M., & Schraefel, M. C. (2006). *Investigating user tolerance for errors in vision-enabled gesture-based interactions.* Paper presented at the AVI '06, Venezia, Italy.

Kelley, J. F. (1984). An iterative design methodology for user friendly natural language office information applications. *ACM Transactions on Office Information Systems*, *2*(1), 26-41.

LaLomia, M. J. (1994). *User acceptance of handwritten recognition accuracy.* Paper presented at the CHI '94.

Li, Y., Welbourne, E., & Landay, J. T. (2006). *Design and experimental analysis of continuous location tracking techniques for Wizard of Oz testing.* Paper presented at the HI 2006, Montreal, Canada.

Lyons, K., Skeels, C., & Starner, T. (2005). *Providing support for mobile calendaring conversations: A Wizard of Oz evaluation of dual—Purpose speech*. Paper presented at the Mobile HCI, Salzburg, Austria.

Oh, A., Fox, H., van Kleek, M., Adler, A., Gajos, K., Morency, L., & Darrell, T. (2002). *Evaluating look-to-talk: a gaze-aware interface in a collaborative environment.* Paper presented at the CHI2002, Minneapolis, MN.

Paiva, A., Andersson, G., Hook, K., Mourão, D., Costa, M., & Martinho, C. (2002). SenToy in FantasyA: Designing an affective sympathetic interface to a computer game. *Personal and Ubiquitous Computing*, *6*, 378-389.

Paive, A., Chaves, R., Piedade, M., Bullock, A., Andersson, G., & Höök, K. (2003). *SenToy: A tangible interface to control the emotions of a synthetic character.* Paper presented at the AAMAS '03.

Pettersson, J. S., & Siponen, J. (2002). *Ozlab—A simple demonstration tool for prototyping interactivity.* Paper presented at the Nordichi 2002, Aarhus, Denmark.

Read, J. C. (2006). *Wizardry and the tangible user interface*. Paper presented at the DCC'06 Workshop—Let's get physical: Tangible interaction and rapid prototyping in, for, and about design, Eindhoven.

Read, J. C., MacFarlane, S. J., & Casey, C. (2003). *Good enough for what? Acceptance of handwriting recognition errors by child users.* Paper presented at the IDC 2003, Preston, UK.

Read, J. C., Mazzone, E., & Höysniemi, J. (2005). *Wizard of Oz evaluations with children—Deception and discovery.* Paper presented at the IDC2005, Boulder, Colorado.

Sinha, A. K., Klemmer, S. R., Chen, J., Landay, J. A., & Chen, C. (2001). *SUEDE: Iterative informal prototyping for speech interfaces.*

Takayama, L., Leung, L., Jiang, X., & Hong, J. I. (2003). *You're getting warmer! How proximity information affects search behavior in physical spaces*. Paper presented at the CHI 2003, Ft. Lauderdale, FL.

Thomas, J. C. (1976). *A method for studying natural language dialog* (Res. Rep. RC 5882). IBM Thomas J. Watson Research Center.

KEY TERMS

Deception: The process by which a user of a system is deliberately made unaware of, or misinformed of, that which is going on.

Recognition-Based Systems: Systems which rely on speech, gesture, or handwriting where the input has to be interpreted by the computer before it can be used by the system.

Robot-of-Oz: A Wizard of Oz system which is fully automated.

Simulation: The use of a mocked-up interface to evaluate user behaviours.

Wizard: The person in a Wizard of Oz study that is doing the simulation. Generally hidden from the subject in the study, the wizard has to 'pretend' to be the computer.

Wizard of Oz: A range of methods in which some or all of the interactivity that would normally be controlled by computer technology is simulated by a human user (known as the wizard).

Wizarding: The process by which a wizard acts to present a simulated response to a user. Sometimes wizarding is supported by a wizard interface.

Wozzing: The process of planning, carrying out, and evaluating a Wizard of Oz study.

Chapter XLVIII
Cognitive Models as Usability Testing Tools

Vanja Kljajevic
NewHeights Software, Canada
Carleton University, Canada

ABSTRACT

This chapter discusses the idea that using computational cognitive models in usability testing has many benefits over the traditional approaches. It argues that computational cognitive models, anchored in the concept of cognitive architecture, offer an integrated approach to interactive behaviour emerging from the use of mobile phones. A cognitive architecture is a theoretical framework containing a set of relatively independent core constraints that are constant across time and tasks. It constrains models built within the cognitive theories based on the architectures, preventing proliferation of implausible theories. This proliferation, on the other hand, is typical of the traditional approaches to usability testing. In this chapter the benefits of using the model-based approach based on a cognitive architecture in usability testing will be discussed, with a special emphasis on mobile phone interfaces.

INTRODUCTION

Over the last 10 years, mobile phones have evolved from simple devices intended for making calls to devices with a rich set of features and many voice and data services. These include: downloadable ring tones, games, short message service (SMS), picture messaging, Internet browsing, e-mail access, and conference calling, and incorporate MP3 player, digital camera, video telephony, television, and information on traffic, mapping, and directions, as well as services for connecting physical and virtual urban life. New features and services that facilitate many activities of everyday life are being added to mobile phones so fast that nothing strikes us as a surprise any more. We can only expect a further increase in functionality of mobile phones, wondering, for example, why turning mobile phones into motion-detection game controllers has not happened earlier.

However, the increased functionality has brought about the complexity problem, opening the question of what the optimal interface of such a complex interactive system would be. Namely, the problem here is to make a single device's interface satisfy users' needs for features and services that were previously provided by several devices. A new device requires a new conception of interface, with an easy access to features and seamless transition among the applications/services. The increased complexity requires more flexibility than the desktop computer-like interfaces of the current mobile phones allow. It is not only that the complexity problem requires a new and more integrated approach to user interface (UI) design of such devices, but it also further complicates the nature of usability testing of these devices, while usability testing is expected to support the UI design processes. It is impossible to empirically assess all the possibilities of a mobile phone UI design by means of traditional usability testing. This chapter claims that computational cognitive models are potential tools for such an assessment. Thus, it is claimed that modeling based on cognitive architectures may be a better alternative to theoretically unsupported, time-consuming, and often expensive traditional usability testing.

The main claims of this chapter are:

a. Usability testing requires solid theoretical underpinnings, the lack of which results in inconsistent and unreliable testing methods.
b. Quantitative usability testing is preferred to qualitative evaluations.
c. Computational cognitive models have the potential to become indispensable usability testing tools.

In order to support these claims, background information is first presented, introducing the fragmentation problem and showing how it is currently reflected in both theory and methodology of usability testing. After introducing the necessary conceptual and terminological clarifications pertaining to the word 'usability,' some of the problems arising from the recently proposed ISO quality models are discussed. The problems indicate that usability testing models need a solid theory, supporting the claim (a). Additional support for this claim comes from the argument on the recently identified evaluator effect found in three most widely exploited usability evaluation methods (UEMs): cognitive walk-through, heuristic evaluation, and thinking-aloud. This section also provides support for the claim (b). The main section of the chapter supports claim (c) and is devoted to usability testing from the computational cognitive modeling perspective and examples of such models pertaining to mobile phones. The rest of the chapter outlines the topics for future research, followed by a brief conclusion.

BACKGROUND

Because of the striking lack of consensus on the definition of usability, it is necessary to make the following distinctions:

1. The general ergonomic definition of usability vs. usability of interactive software applications
2. Usability of a software product vs. its quality in use

From a general ergonomic point of view, usability of a product is its fitness for purpose. The definition actually refers to the product's functionality, although it also implies that a product is usable if it can easily facilitate user's completion of a task at hand. When applied to a user interface, which most users usually identify with the whole product (Mayhew, 1999), the general definition of usability fails, because it is too vague. The main question that an attempt to apply the general ergonomic definition of usability to a user interface is: How do we determine a product's usability, given that what all the users notice is the interface, while, on the other side, behaviour and appearance of the elements of the interface depend on the system's underlying structure and code (Zetie, 1995; Schneiderman & Plaisant, 2005)? In addressing this question, the International Organization for Standards (ISO) has published several standards and standard type documents on software usability, placing the concept within carefully designed quality models. The models suggest that usability is an aspect of software quality, postulating different usability components within different versions of the quality model. The question also indicates the need to distinguish between the internal software qualities and its external characteristics, that is, qualities that *users perceive*. These internal qualities influence users' impressions on the product's quality. In order to consistently distinguish between the two types of quality characteristics, the ISO has introduced the term *quality in use* for the latter type.

Unlike usability, quality in use is not a software quality characteristic in the strict sense, but rather a user's impression of the quality of the system containing software (ISO 9126-4). Given that quality in use can be influenced by any of the software quality characteristics, including usability, and since it emerges from the combination of these characteristics, the concept of quality in use is wider than the concept of usability. Thus, the questions such as Lewis's (2001) on how 'likeability' or 'appealingness' relate to usability actually point to the quality in use. We can begin to measure these attributes by first adopting the distinction between usability and quality in use.

Although usability is a key concept of UI design, it is not always that well integrated into the design process (Hartwig, Darolti, & Herczeg, 2003; St. Amant, Horton, & Ritter, 2006). Even when it is a part of UI design, usability testing usually boils down to only one usability evaluation per product (Lewis, 2001). This seriously limits the potential of usability intervention on an interface improvement, making a remarkable 60% of interactive software defects the usability issues (Bevan, 1999).

Ideally, usability testing is part of user interface design, where its function is to find the problems. In other words, its function is not to collect quantitative data per se, but rather to use the collected data as indicators of what needs to be improved in order for a product to achieve a quantifiable usability level determined for a specific user group and a specific task. Other functions of usability testing are to enable comparison of products or to measure a product against the standards/guidelines (Lee, 1999).

Both the methodology of and theorizing on usability testing are marked by profound fragmentation. The fragmentation in methodology of usability testing is caused by the lack of an adequate theory and is evident in formal and informal usability testing. For example, those who conduct formal usability testing often concentrate on a small set of tasks, facing the questions such as: How to avoid the bias in task selection (Lewis, 2001)? How to ensure the comprehensiveness of testing and the replicability of results? How to provide correct interpretation of the severity of errors, make the right choice of metrics and ensure their reliability, as well as how to correctly predict whether the usability determined in a lab will hold in the 'real world'? Similarly, those who perform informal usability evaluation face the questions of reliability and validity of the methods they use (Hertzum & Jacobsen, 2001).

These open questions limit the potential contribution of usability testing to interface design, indicating the need for a comprehensive theory underlying the testing. Namely, most of the research on user interface design still heavily relies on micro-theories developed within a certain paradigm in psychology, addressing only a specific aspect of cognition or perception (e.g., situational awareness, visual processing, attention, working memory, etc.). This reduces what should be an explanation of the rich dynamics of interactive behaviour to only those aspects of human behaviour that are interesting to a particular researcher at a particular time, making them random, theoretically unmotivated choices. Since interactive behaviour emerges from the dynamic relations between the user, a task, and an artifact, it is crucial that usability testing capture the totality of this triad's dynamics. In addition, interactive behaviour engages each individual user as a unique complex of cognitive and perceptual processes and motor activities. Thus, because they focus on a single element of cognition or perception (e.g., spatial working memory, visual attention), or at best on a single relation within the user-task-artifact triad (e.g., cognition-task), disregarding the other two elements (task, interface) and other relationships within it (Gray & Altmann, 2001), these explanations are too fragmented to provide a coherent overall picture of *interactive* behaviour and contribute to user-interface design (Byrne, 2001).

This problem is particularly emphasized by the complexity problem, which is becoming increasingly evident in mobile phones and more pronounced in their use on the move. Namely, multitasking behaviour typical for mobile situations (e.g., driving and dialing a number on a cell phone) imposes strategies for setting priorities among the tasks as well as trade-off between the cognitive resources allocated to each of them. The mobile and the mobility tasks compete for the limited cognitive resources. This causes depletion of resources for some tasks, leading to the breakdown of fluent interaction. Thus, research and theoretical insights into any specific element of mobile interaction (e.g., deployment of visual gaze, motor control, attention) are important for our *understanding* of this type of psycho-social behaviour. However, it is precisely their focus on a single/specific aspect of the multitasking mobile behaviour that makes such studies too specific to *test the totality of this behaviour*—and testing the totality of mobile behaviour is a prerequisite if usability testing intends to improve UI design of mobile devices.

In contrast to this fragmentation, cognitive science offers a unified approach to interactive behaviour in mobile contexts. The approach is anchored in the concept of cognitive architecture. A *cognitive architecture* is a theoretical framework whose hallmark is a specific set of relatively independent core constraints that are constant across time and independent of tasks. An architecture constrains cognitive *theories* developed within it, preventing proliferation of implausible theories. A cognitive theory further constrains a *model*, which is usually a running computer program that produces real-time human-like behaviour in an interactive setting (Ritter et al., 2006).

There are two types of reasons for a wider use of computational cognitive models in usability testing. First, these models are theory-based. The theoretical framework of an architecture provides coherent principles for the overall organization of the theories and models within it (supporting (a)). In addition, the computational models based on cognitive architectures are capable of providing quantitative answers to usability questions (supporting (b)). The literature abounds with examples of static check-lists for qualitative usability assessments (for example, see Ji, Park, Lee, & Yun, 2006, for such a checklist for the usability of mobile phone user interfaces). These assessments cannot encompass all the aspects of usability testing, because usability is not an absolute, but rather a graded category, within which small degrees of deviation from the required and expected may represent a subtle change in design, but would require larger execution times, leading to a huge financial loss in the long run. Thus, the graded nature of usability is better captured by detailed, quantitative specifications than by all-or-none qualitative evaluations.

However, although quantification is preferred in usability testing, if not theoretically supported, it is insufficient. The efforts of the International Organization for Standards to establish a principled quantitative approach to usability testing and determine a reliable set of coherent metrics illustrate the point (see the second section). Similarly, there exist psychological models that provide accurate quantitative predictions but lack the theoretical apparatus needed for analyzing the causes of the behaviour they model. Both types of models lack the theoretical power and constraints of a framework such as a cognitive architecture. From this point of view, computational cognitive models have advantage over the models lacking adequate supporting theories.

The second set of reasons for employing computational cognitive models in usability testing is circumstantial. Unlike the substantial reasons, such as the theoretical grounding of the models and their capability to consistently capture and explain the elements and relations within the user-task-artifact triad, the circumstantial reasons are imposed by the circumstances in which usability testing is being conducted. They include prohibitive costs, impracticality of conducting testing (e.g., conducting mobile phone interface experiments with a user while he/she is driving), or lack of time for conducting formal usability testing. Computational cognitive models have the means to successfully overcome the problematic circumstances, providing tools for fast and reliable quantitative testing of a large range of design possibilities.

USEFUL AND USABLE?

In a wider sense, testing is any examination or evaluation of quality or functionality of something. A more specific definition would point out that testing is the use of empirical tests for the purpose of obtaining quantified data on certain aspects of a particular phenomenon under specific conditions. Testing may also mean the measurement of specific qualities of a phenomenon and comparing them to standards. When it comes to usability, the different meanings of 'testing' are all reflected in the divergence of methodology: There exist informal usability evaluation methods, formal usability testing, standardized usability measures, and recently computational cognitive models have begun to be applied to aspects of usability testing. Not all of these methods are useful and/or usable. The next section briefly discusses measures proposed by the ISO quality models, illustrating the point that models without adequate theories at best provide only partial insights into a product's usability.

Models without Adequate Theories: The ISO Quality Models

It has been repeatedly emphasized in the literature that the end-point of user-centered design is not a *product*, but rather *a set of interactive behaviours* between a user and an interface (Bevan, 1999; Byrne, 2001; Grey & Atmann, 2001). For example, the standard *ISO/IEC 9126* (1991) postulates that software quality emerges as a combined effect of usability, functionality, efficiency, reliability, portability, and maintainability—when a product is *in use*. Although the model acknowledges that the design should meet users' needs, its focus on a product's external qualities rather than on interactive user-interface behaviour makes it a product-centered model. Similarly, *ISO 8402* (1994) proposes that usability of interactive software is a set of features that can be *designed into the product*, implying that it is possible to achieve usability without a user. Another relevant standard, *ISO 9241*-11 (1998), claims that usability determinants are effectiveness, efficiency, and user's satisfaction with software. However, these determinants of usability also depend on the other five software quality characteristics listed, and in addition—because they are abstract characteristics—they cannot be measured directly (Bevan & Macleod, 1994).

ISO/IEC 9126: 1-4 (2003) has proposed a new, generic chain quality model, according to which software quality consists of the following elements:

1. Internal quality (properties of the code), which impacts
2. External quality (software's behaviour when it is in use), which in turn has impact on
3. Quality in use (the extent to which software meets the needs of the user).

Usability in this model is still an abstract characteristic, defined as ease, with which software can be understood, learned, used, and is attractive to the user. Thus, usability is still an abstract characteristic of software quality that can be measured only indirectly, via measurement of the attributes of its sub-characteristics. Conceptually overlapping in this model are product quality and users' needs. It is interesting that the metrics system fails to reflect their conceptual overlap within the model, specifying the sets of matrices for measurement of the internal and the external aspects of usability, and the quality in use. The problem here is not the generality of the proposed metrics per se, but rather the fact that the measurements they enable can only be descriptive and done through interpretation, because the quality characteristics are expressed in natural, and not in a formal language (logic, mathematics) (King, 2003). Given that the purpose of a quality model is to guide software development and enable its objective evaluation and measurement, the proposed model seems to fall short of such a goal.

As King (2003, 2006) has noticed, these models have not determined yet what makes good metrics, let alone presented a consistent framework for implementing a measurement program. The ISO metrics have been criticized in the literature as incoherent, unclear, non-intuitive, and expensive, with weak or no correlation with human judgment, and so forth (King 2003, 2006). Note that the metrics proposed by the ISO are without theoretical constraints, which means that one could in principle keep adding new metrics indefinitely. King's recommendation to constrain the metrics according to criteria such as reliability, intuitiveness, correlation with human judgments, and so forth, is a way to improve the situation, which could be even more strengthened by strictly imposed, theory-inherent constraints.

That King's criticism holds is illustrated by the recent introduction of two new, metrics-related concepts in the quality model in the 'next generation' of ISO standards, *SQuaRE*. The new concepts are *quality measures* and *measurement primitives*. It is not clear how introduction of the new concepts improves the quality model (see Abran et al., 2005 for a review). Rather, it indicates that the proliferation of new terms and components within the quality model only reflects the lack of a coherent underlying theory that would constrain the models, resulting in consistent, reliable, and more usable metrics.

Thus, given the complexity problem and the nature of multitasking behaviour in the use of mobile phones, on the one side, and the lack of a solid theoretical basis for its current quality models, on the other side, the current ISO proposals do not seem to provide a useful and usable model for usability testing of mobile phone user interfaces.

Some Problems with the Qualitative Methods

Usability testing often refers to any type of usability evaluation, regardless of the type of evaluation technique (Rubin, 1994). However, some researchers differentiate between formal and informal usability evaluation methods. The former is a process of obtaining empirical evidence in a controlled, laboratory setting from the representatives of a target users' group on tasks specifically designed to assess different aspects of a product's usability. In this setting, an informal technique known as thinking-aloud is also used as a source of evidence for potential usability problems. The participants performing tasks are required to 'think aloud' so that observers and the experimenter can collect additional data about the interaction between a user and an interface/device. The method requires teaching the participants how to think aloud and encouraging them to do that while solving tasks on a system that is being tested. Not only are participants often not comfortable with this method, which questions its acceptability, but also the experimenter's intrusions in the form of 'encouragement' of participants to think aloud question the validity of the attempted formal testing.

Other widely used informal UEMs are heuristic evaluation and cognitive walkthrough. Heuristic evaluation is a process in which a set of principles guides a group of evaluators in judging the compliance of an interface with the principles. The major problem with this method is that it is based on heuristics. The number of heuristics is in principle indefinite, with nothing setting apart the 10 usually applied heuristics (Nielsen, 1994) as more basic than other heuristics that one could come up with. In cognitive walkthrough, the evaluators' task is to imagine users' performance on certain tasks involving the interface. The evaluators are required to decide whether a 'typical user' would manage to perform the correct sequence of steps required for a specific task. This method, like other informal UEMs, leaves plenty of room for personal judgment.

In addition, it is not clear whether an evaluator can actually imagine the impact of required perceptual sampling in different environments (crowded bus, highway, café) on interaction with the mobile interface. Or can he/she imagine the impact of the interplay between the higher-level cognitive functions (long-term memory, inferencing, thinking, decision making) and environment monitoring on the use of a mobile phone? Although an evaluator might perhaps correctly predict task sequences in some cases involving non-mobile environments, the processes involved in mobile behaviour are too complex to be left to an evaluator's imagination. The recent finding on "the almost eight-fold differences between the lab, our baseline situation, and the street situations" in experiments on attentional resources in mobile phones illustrates the point (Oulasvirta, Tamminen, Roto, & Kuorelahti, 2005).

It is not surprising then that Hertzum & Jacobsen (2001) have recently shown that cognitive walkthrough, heuristic evaluation, and thinking-aloud exhibit the so-called *evaluator effect*. Namely, Hertzum and Jacobsen analyzed the empirical data from 11 studies based on these three UEMs and found that the multiple evaluators who evaluated the *same interface* by the *same methods* most often reported *different* usability problems. When they did agree on the existence of a particular problem, they ranked its severity differently. Also, the evidence indicates that the methods did not show whether the problems reported by an evaluator also existed in use outside the usability evaluation setting. The empirical data presented by Hertzum & Jacobsen show that due to the evaluator effect, cognitive walkthrough, heuristic evaluation, and thinking-aloud are neither reliable nor valid usability evaluation methods.

Gray & Salzman (1998) have also questioned the reliability and validity of UEMs, pointing out to potentially misleading messages stemming from some prominent usability studies. The main problems with these studies, according to Gray & Salzman, are that first of all they treated usability as a "monolithic, atheoretical construct," and second, the experimental design in these studies was inadequate with regard to different types of validity. Both objections indicate a need for solid theoretical underpinnings and a stronger role of theory in usability testing, experimental design, and interpretation of results.

So far, we have seen that the lack of solid theoretical grounds that would impose theory inherent constraints on ISO metrics, on the one hand, and the evaluator effect found in widely applied UEMs such as cognitive walkthrough, heuristic evaluation, and thinking-aloud, on the other, present serious obstacles to accepting either approach as the optimal way of usability testing of mobile phone interfaces. In the next section, the possibility of using computational cognitive models for usability testing is turned to.

COGNITIVE ARCHITECTURES

This section develops arguments for the third claim of the chapter:

(c) Computational cognitive models have the potential to become indispensable usability testing tools.

A cognitive architecture is a theoretical framework intended to provide a unified explanation of those aspects of human information processing that are relatively invariant over time and across tasks. Cognitive architectures are often contrasted with expert systems. The latter are problem-solving computer programs limited to problems in a specific field that require significant human expertise (Wagman, 1997). The former are broad theories on human behaviour, within which descriptive and generative (i.e., computational) models of cognition, perception, and action have been built, with a promise of successfully incorporating even moderators' variables, such as individual differences in personality, cultural background, stress, fatigue, and so forth (John & Altmann, 1999; Sun et al., 2004). Thus, cognitive architectures integrate theories on human cognition, perception, and motor actions together with theories on how these behaviours occur and function in a specific context under specific circumstances.

Newell's (1972) original idea on unification of psychological research into a broad, generic theory of cognition, which would explain 'the full range of human behaviour,' has been extended to encompass socio-cultural contexts in which behaviour takes place, as well as other analyzable levels of behaviour, all the way to its neural underpinnings (Sun et al., 2004; Anderson et al., 2004). Thus, cognitive architectures intend to comprehend and account for cognition and behaviour 'at the systems level' (Langley, Laird, & Rogers, 2006), using modeling and simulation as methodology fit for these different levels of analysis and their unification. Models simulate human behaviour (e.g., duration of an action, occurrence of error in a task, learning strategies, learning rates, transfer of knowledge, etc.), predicting for example, types and rates of errors in a certain task, learning time for specific actions, and so forth. In addition to their predictive power, models have the capabilities to explicate and test theoretical assumptions of their architecture, refining the architecture by prohibiting inconsistent assumptions. In order to run as computational programs, models require theoretical specifications at the level of fine details, demanding consistent and coherent theoretical explanations from their architectures. This approach clearly differs from the traditional research in psychology, which prefers specific theories on specific components of human cognition and behaviour, with separate methods for 'specialized tasks' (Byrne, 2002; Sun, 2004; Sun et al., 2004; Langley, Laird, & Rogers, 2006). As pointed out by Newell (1973), accumulation of data through fragmented theories does not lead to progress in science, although it may leave that impression. More importantly, they might mislead design (Gray & Salzman 1998).

Since cognitive architectures intend to account for the full range of human cognition and behaviour, they need to be capable of performing many functions and thus need to be equipped with many functional capabilities. The cognitive architectures that have been built so far support, either directly or indirectly, a wide range of activities, such as categorization, decision making, problem solving, perception, recognition, prediction making, monitoring, planning, interaction, learning, and so forth. These functional capabilities are enabled by the internal properties of cognitive architectures, such as knowledge representation, utilization, and acquisition (Langley, Laird, & Rogers 2006).

Cognitive architectures are roughly divided into descriptive, such as GOMS (goals, operators, methods, and selection rules) (Card, Moran, & Newell 1980) and generative architectures, such as Soar, EPIC, and ACT-R. They differ in theorizing on the mind and impose different types of constraints on their models. All the models built within one architecture share the same features, but in addition they all have their respective, model-specific, extra information—*the model increment* (Howes & Young, 1997). Cognitive architectures are theoretically well defined, and as such they impose constraints on the model increments to either facilitate certain types of model constructions within the architecture (the soft constraints) or to enforce some and rule out other features (the hard constraints).

The models based on cognitive architectures have been used in human-computer interaction in three ways: to predict task performance times and errors, to assist users in task performance, or to substitute users in testing (John, 1998; Ritter et al., 2000). Given the theoretical basis provided by cognitive architectures and the constraints they impose on the models, as well as the additional knowledge in the form of model increment, it is reasonable to expect that these models would better contribute to UI design than the user models which are directly implemented in a general purpose programming language without such an architectural anchor, for example, Lisp or C (Howes & Young, 1999). The latter type of user model lacks the mechanisms for employing the insights on user's mental models in design and cannot correctly predict user's behaviour. As a consequence, these models are testing a modeler's predictions, and not the predictions made by a model itself. That is not the case with the models based on cognitive architectures. Furthermore, researchers have recently added a simulated eye and a simulated hand to some computational cognitive models (Ritter et al., 2000) and have begun adding 'overlays' to the models in the architectures in order to enable modeling of cognitive moderators (e.g., stress, fatigue, anxiety and other emotions, personality, etc.) (Ritter, Klein, Quigley, & Schoelles, 2004; Gluck, 2004). Adding a simulated eye and a simulated hand enables modeling of visual perception and motor actions which are the key elements of interactive behaviour required by the current user interfaces.

Adding behaviour moderators to computational cognitive models has a great potential for modeling users in usability testing because it widens the scope of the testing conditions and tasks with simulated users. This is another reason for which computational cognitive models may be preferred to the traditional usability testing methods, because conducting experiments with certain behaviour moderators is impractical or even dangerous. Adding behaviour moderators is crucial in user modeling intended for mobile phones, because interaction on the move imposes huge demands on actual users' ability to multitask, resulting in a variety of responses that depend on factors that none of the currently used methods can successfully test or measure. In addition, the architectural models enabled to model cognitive moderators have the potential to improve predictions on and testing of the *quality in use*. This can be done by first collecting data on a small number of actual users' performance on specific tasks in a pilot study, and then using the data to create a user profile (cf. St. Amant et al., 2006). The users' models implemented in general purpose programming languages such as Lisp or C lack this potential.

Descriptive vs. Generative Cognitive Models

The optimal UI design maximizes the effect of cognitive processes implicated in a particular task while minimizing the cognitive effort that these processes impose. This is usually achieved through the iterative development of models, which reflect understanding of user's mental states and the models involved in the task. In order to make use of the mental models, a designer needs to find a mechanism for adequate external representation of the mental states. In other words, he/she needs to build a cognitive model and simulate mental task processes. A cognitive model is an external representation of mental states. The activity of mapping of the mental models onto cognitive models is known as cognitive modeling (Peschl & Stary, 1998). In computational cognitive models simulation of mental task processes is performed in a computerized model. Thus, since cognitive models developed within cognitive architectures are often generative, that is programmable—they can run on a computer—there is a tendency in the current literature to use the term 'cognitive model' exclusively for computational cognitive models. This is misleading because there exist descriptive cognitive models, which also successfully model cognitive aspects of interactive behaviour. GOMS is an example

The GOMS concept (John & Kieras, 1996b) proposes that the knowledge of a task should be analyzed in terms of *goals*, *operators*, *methods*, and *selection rules* (Card, Moran, & Newell, 1980). *Goals* refer to the users' goals. *Operators* are users' actions licensed by software. For example, while the command-line interfaces would have a command and its parameter as an operator, in graphic user interfaces (GUIs) operators are menu selections or button presses. In addition, gestures and voice are becoming increasingly important operators, and eye-movements are expected to become that soon, too. *Methods* are sequences of sub-goals and operators used to achieve a goal. In a case when there are several methods to accomplish a specific goal, a user will employ *selection rules* to decide which method to choose.

Several variants of the GOMS models have been used as user interface analysis techniques so far (John & Kieras, 1996a):

- The original Card, Moran, and Newell version of GOMS, also known as CMN-GOMS (where CMN refers to Card, Moran, & Newell's 1980s model)
- The Keystroke-Level Model (KLM), which is the simplest version of the model
- Natural GOMS Language (NGOMSL)
- Cognitive-Perceptual-Motor GOMS (CPM-GOMS)

What they all have in common is making quantitative and qualitative predictions of users' behaviour, such as estimates of task performance time.

Project Ernestine (Gray, John, & Atwood, 1992) is perhaps the most famous example of a cognitive model based on a GOMS technique that saved millions of dollars of a company's money. The project was intended to determine whether introducing new workstations would indeed save 4 seconds per average phone call, adding up to savings of 12 million/year, as the manufacturer of these workstations had suggested. The scientists running *Project Ernestine* used CPM-GOMS to analyze cognitive, perceptual, and motor operators and dependencies among them in the tasks that toll and assistance operators (TAOs) performed in operator-assisted calls. The cognitive model predicted that the new workstations would have actually added 1 second to the average call, and the increase in time would have increased the operating costs of the company by an additional $2.4 million (Gray, John, & Atwood, 1992). The predictions of the model were confirmed by the empirical field data.

This is an example of the *predictive power* of a cognitive model. Note, however, that the model also provides an *explanation* for the increased time/costs that the new workstations would have imposed. Namely, the user model that the manufacturers employed was obviously counting only on the reduction of the number of steps in the task, that is, the fact that the new workstations would have reduced the number of the required keystrokes in call processing. However, this model disregarded the impact of the context of the call (e.g., simultaneous cognitive, perceptual, and motor actions). On the other hand, the cognitive model based on CPM-GOMS considered the elements of context, correctly predicting more steps with the new workstations and therefore a bigger financial cost.

There are other examples illustrating that the GOMS techniques can also make useful quantitative predictions on user's learning of the procedure that he/she must follow in order to complete a task, capturing procedural speed and complexity (John & Kieras, 1996). In a recent study, Luo & John (2005) explored whether KLM (aided by CogTool consisting of ACT-R cognitive architecture, Macromedia Dreamweaver, and some other components) can accurately predict task execution time on handheld devices and found less than 8% error. In another study, Salvucci & Macuga (2001) developed an integrated computational model, based on four models implemented in the ACT-R cognitive architecture, to predict the impact of a cell phone dialing on driving. The model predicted, and the empirical data confirmed, the larger effects of dialing methods with 'higher visual demands' than those with 'long dialing times.' Although it has been pointed out in the literature that the cognitive architectures underlying GOMS models are too simple to capture *all* the aspects of human performance in an interactive task, recently St. Amant, Horton, & Ritter (2004) have shown that a GOMS-based model was "slightly better" than an ACT-R based model in evaluating mobile phone menu hierarchies.

In order to fully comprehend the meaning of this finding we need to briefly introduce ACT-R (Atomic Component of Thought with Rational Analysis). ACT-R is a powerful generative cognitive architecture that has been evolving since the 1970s. In the 1990s the perceptual and motor components were added to that model as ACT-R/PM (PM referring to perceptual and motor), and the current versions are ACT-R 5.0 and ACT-R 6.0 (Anderson et al., 2004). Recently, an attempt to make the cognitive architecture encompass the neural underpinnings of human behaviour has been made and ACT-R's modules such as visual, declarative, episodic, manual, a goal module and so forth, which are assumed to correspond to the human cognitive modules, as well as other elements of the model, have been mapped onto the brain structures (Anderson et al., 2004). ACT-R 5.0 has been successfully applied to many domains of human cognition, perception, and motor action. This version of the model was used in St Amant et al.'s (2004) study on menu traversal on a mobile phone.

Namely, St. Amant et al. conducted an empirical study on menu traversal on a mobile phone. Designing menu hierarchy for mobile phones is a challenging task not only because these devices have small screens that are usually indirectly operated (the lack of a mouse or touch screen requires a discrete selection of actions for moving from one menu item to another), but also because of the lack of appropriate standardization of hardware supporting the menu traversal on these phones (St. Amant et al., 2004). Thus, having a model that accurately predicts a user's performance on the menu traversal on mobile phones would help usability testing, and a search procedure that would generate menu hierarchies with reduced traversal time would improve their UI design.

The study collected data on performance of 12 actual users and compared them with the predictions on users' performance that three models of different levels of abstraction made on the same five menu traversal tasks. The three models are: a Fitt's law model, a GOMS model, and an ACT-R 5.0 model. They were developed on the basis of data obtained from a single task performance of a single user in using mobile phone menus in a pilot study. The authors found that, although many models of the interactive mobile phones behaviour have been based on Fitt's law (e.g., models testing keypad dialing and one-finger text entry), their Fitt's law model was inferior to the other two models in predicting users' behaviour. Note that Fitt's law models are not based on a cognitive architecture and are not equipped to deal with cognitive and visual processing, which are implicated in actions that need to be performed in mobile phones menu traversal tasks. It is not surprising then that the predictions of the Fitt's law model were less accurate than the predictions of the other two models, both of which gain explanatory power from their respective cognitive architectures.

However, since ACT-R is a *generative*, far more developed and powerful cognitive architecture than GOMS, which is a coarse, *descriptive* architecture, it is surprising that the GOMS model was slightly better in making predictions than the ACT-R 5.0 model. For example, the latter incorrectly predicted that scrolling actions take more time than selection actions. Given that GOMS models are cheaper to build, it makes sense to consider applying coarser formalism when modeling simpler procedures. Also, the coarser architecture of the GOMS model turned out to be more flexible, allowing for example a specific ordering of decision steps to be not governed by the cognitive constraints, but rather modeled on a single user's behaviour, which was in this case successfully generalized to a larger sample. This, however, is not the case with the ACT-R 5.0 model, whose architecture is less abstract and therefore imposes tighter constraints on the model. Like the CAPS model, which will be presented, these two models indicate that the depth of menu hierarchies poses a problem to mobile phone users. In a similar study, St. Amant et al. (2006) found that menu efficiency can be improved up to 30% when the menu hierarchy depth was reduced and the individual menus length was increased. However, this finding is in conflict with both behavioural and some earlier modeling data indicating that, contrary to our intuitions, the error rates do not decline with the decrease in number of phone menu items. This further indicates that the shorter menus 'are not necessarily better' (Byrne, 2002). An architecture that makes correct predictions and presents an explanation of this fact is CAPS.

Namely, CAPS (collaborative activation-based production system) (Just & Carpenter, 1992) is an interesting programmable cognitive architecture whose unique theoretical quality comes from enabling modeling of individual differences in working memory (WM) capacity and the impact that WM load has on a task performance. Phone-based interfaces create excessive WM load: Users have to *remember* the presented options. An interesting problem they pose is whether to allow deeper menu hierarchies with more items per menu, or to restrict the menu items to a smaller number of general items, preventing deep hierarchies. Deep hierarchies pose WM load which represents an obstacle for successful processing, especially in mobile situations. The theory underlying the CAPS computational model explains that the increase in the number of menu items and the resulting depth of the menu hierarchy create a higher working memory load, requiring more WM than is available, thus outstripping the available resources. On the other hand, longer menus also create higher WM load, because they contain more general items; and the more general the item, the more resources are required for its disambiguation. This explains the finding discussed, indicating that the error rates do not decline with the decrease in number of phone menu items and with the reduction of the hierarchy's depth (Byrne, 2002).

These examples show that cognitive models and cognitive architectures have been successfully applied to several domains of UI design and usability testing of mobile devices. It is surprising and disappointing that they are not more often present in usability testing and UI design of mobile phones. Namely, there are models that are so general that they can be used for different traversal tasks, for different menu hierarchies, and on different mobile phones—without requiring modeling background or additional programming skills from a developer (St. Amant et al., 2004). Given the importance of usability in iterative design, ignoring such models only indicates a lack of communication on relevant findings among the fields participating in UI design and testing.

FUTURE TRENDS

Specification of the Existing Architectures

Further development of cognitive architectures is inevitable, not only through refinements of the existing architectures, but also by the development of new ones. The driving force of this development is the need to understand the interactive behaviour and at the same time to strengthen the ties among the disciplines that participate in UI design, providing powerful modeling and usability testing tools. This development is related to the second major future trend in the field development of an integrative approach to UI design.

An Integrative Approach to UI Design

An integrative approach to UI design is an approach based on the principles that unify cognition and other elements of interactive behaviour. These principles also unify theories on interactive behaviour and their methodology with UI design. For example, the principles of economy, information structuring, and relevance are such unifying principles. The current state of affairs in usability testing indicates the need for a theoretical framework that would capture the dynamics of interactive behaviour within the user-task-artifact triad in a principled way. Since cognitive architectures provide such a framework, it is reasonable to expect their wider presence in UI design and testing in the future. Thus, further development of theory and methodology of usability testing requires consistent application of the unifying principles in an approach based on a cognitive architecture.

Promoting Flexible Design

Unlike traditional phones, which are typically used by more than one person, mobile devices usually belong to one person. Such a use requires a more *flexible* design, allowing faster and more effective personalization of the device and its applications. Since mobile devices are used on the go, personalization that is optimal for one situation may turn out to be nonfunctional in another context (e.g., change in visibility requires changes in display). Thus, fast and seamless transition among the variants of personalized applications directly impacts a mobile phone's quality in use.

However, the full meaning of flexible design becomes clear in light of the fact that mobile devices have a unique capability of projecting on end users an impression of total independence. It has gone so far that end users without background in programming now want to develop applications for their devices. This is a big challenge for those who are developing environments for development of such applications. At the same time it represents a great opportunity for those who are developing computational cognitive models as testing tools. Namely, fast and reliable testing tools based on computational cognitive models that have the potential to suggest and check a huge number of different possibilities for UI (the 'what if...' questions, Ritter et al., 2000) may turn out to be even more critical in these cases than in professional UI design settings, in which one may choose other ways to evaluate an application. However, in order for usability testing based on computational cognitive models to become readily available, reusable templates need to be created.

From Usability to Reusability

Reusability is a feature that is desirable both at the level of cognitive architecture and at the level of its models. It is not clear at the moment how to achieve reusability of the content of cognitive architectures. One way might be to have architectures support specific reusable capabilities. Another way would be to perhaps develop reusable knowledge (Langley et al., 2005).

At the level of computational cognitive modeling of interactive behaviour, reusability means building reusable templates of human behaviour (Matessa, Vera, John, Remington, & Freed, 2002). The templates are compiled data on human cognitive, perceptual, and motor performance. There are several benefits of using the templates in cognitive modeling. For example, the templates are based on scientific theories on human cognition, perception, and motor performance. It is also cheaper to build a model with templates than to start from a scratch. Reusing templates is at the same time an excellent opportunity to test the component models and see if the model predicts behaviour on a certain task equally well in a different context. This in turn provides further constraints on an architecture. Finally, because of the rapid changes and increasing demands for quick results in the field, reusable templates may present the only way to make computational modeling of interactive behaviour more present in mobile phones UI design and testing. Implementing reusable templates is an optimal methodological solution for unification of the principles of economy, information structuring, and relevance at the levels of cognition, theory, methodology, and design. Thus, reusable templates are the first step towards routine application of architecturally based cognitive models in UI design and testing.

CONCLUSION

At the moment, usability testing is not a universally practised part of mobile phones UI design. Although the reasons for the hesitance to systematically employ usability testing in UI design of mobile phones are usually presented as circumstantial (e.g., lack of time or money), a closer look at the tools currently available for usability testing reveals that the actual reasons are more substantial. Namely, there is a profound lack of theoretical and methodological foundations for effective usability testing. This chapter has pointed to some problems in traditional usability evaluation and measurement methods that result from the lack of solid theoretical constraints. As an alternative, the idea of computational cognitive models as usability testing tools was introduced.

Although computational cognitive models have the potential to become indispensable usability testing tools, at the moment they are not readily available for all aspects of usability testing. In addition, proficiency in a cognitive architecture requires a significant amount of time and effort. While progress in this area is inevitable, we can begin now to apply the models that have been successfully applied to certain aspects of mobile phone UI design and usability testing. In the meantime, it is important to keep in mind that what makes computational cognitive models good candidates for UI design and usability testing are their solid theoretical grounds and their reusability potential.

REFERENCES

ISO Documents

ISO/IEC 9241-11. (1998). Ergonomic requirements for office work with visual display terminals (VDTs)—Part 11: Guidance on usability.

ISO 13407. (1999). Human-centered design process for interactive systems.

ISO 14598-1. (1999). Information technology—Software product evaluation.

ISO 15939. (2002). Software engineering—Software measurement process.

ISO/IEC TR 9126-1. Software Engineering—Product Quality—Part 1: Quality Model.

ISO/IEC TR 9126-2. (2003). Software Engineering—Product Quality—Part 2: External Metrics.

ISO/IEC TR 9126-3. (2003). Software Engineering—Product Quality—Part 2: Internal Metrics.

ISO 9126-4 9126-4. (2003). Software Engineering—Product Quality—Part 4: Quality in Use Metrics.

Other References

Abran, A., Desharnais, J. M., Qutaish, R. A., & Habra, N. (2005, May). *An information model for software quality measurement with ISO standards.* Paper presented at the International Conference on Software Development, Reykjavik, Iceland.

Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., & Qin, Y.. (2004). An integrated theory of the mind. *Psychological Review, 111*(4), 1036-1060.

Best, B. J., & Lebiere, C. (2006). Cognitive agents interacting in real and virtual worlds. In R. Sun (Ed.), *Cognition and multi-agent interaction: From cognitive modeling to social simulation* (pp. 186-218). New York: Cambridge University Press.

Bevan, N. (1999). Quality in use: Meeting user needs for quality. *Journal of System and Software, 49*(1), 89-96.

Bevan, N. & Macleod, . (1994). Usability measurement in context. *Behaviour and Information Technology, 13,* 132-145.

Byrne, M. D. (2001). ACT-R and menu selection: Applying a cognitive architecture to HCI. *International Journal of Human-Computer Interaction 55.*

Byrne, M. D. (2002). Cognitive architecture. *Cognitive Science, 21*(1), 31-61.

Byrne, M. D., & Gray, W. D. (2003). Returning human factors to an engineering discipline: Expanding the science base through a new generation of quantitative methods. [Preface to the Special Edition]. *Human Factors, 5*(1), 1-4.

Gray, W. D., & Altmann, E. A. (2001). Cognitive modeling and human-computer interaction. In W. Karwowski (Ed.), *International encyclopedia of ergonomics and human factors.* New York: Taylor & Francis, Ltd.

Gray, W. D., John, B. E., & Atwood, M. E. (1992, May). The precies of project Ernestine or an overview of a validation for GOMS. In *Proceedings of the Computer-Human Interaction '92* (pp. 307-312).

Gray, W. D., & Salzman, M. C. (1998). Damaged merchandise? A review of experiments that compare usability evaluation methods. *Human-Computer Interaction, 13*(3), 203-261.

Hartwig, R., Darolti, C., & Herczeg, M. (2003). Lightweight Usability Engineering Scaling Usability-Evaluation to a Minimum? In J. Jacko, & C. Stephanides (Eds), *Human computer interaction—theory and practice* (Part 1, pp. 474-478). London: Lawrence Erlbaum Associate Publishers.

Howes, A., & Young. R. M.(1997). The role of cognitive architecture in modeling the user: Soar's learning mechanisms. *Human-Computer Interaction, 12,* 311-343.

Ji, Y. G., Park, J. H., Lee, C., & Yun, M. H. (2006). A usability checklist for the usability evaluation of mobile phone user interface. *International Journal of Human-Computer Interaction, 20*(3), 207-231.

John, B. E., & Altmann, E. A. (1999). The power and constraint provided by an integrative cognitive architecture. In *Proceedings of the Second International Conference on Cognitive Science and the 16th Annual Meeting of the Japanese Cognitive Science Society Joint Conference,* Tokyo, Japan, (pp. 20-25).

John, B. E., & Kieras, D. E. (1996a). The GOMS family of analysis techniques: Comparison and contrast. *ACM Transactions on Computer-Human Interaction, 3*(4), 320-351.

John, B. E., & Kieras, D. E. (1996b). Using GOMS for user interface design and evaluation: Which techniques? *ACM Transactions on Computer-Human Interaction, 3*(4), 287-319.

King, M. (2003, April). *Living up to standards.* Paper presented at the 11th Conference of the European Chapter of the Association for Computational Linguistics, Budapest, Hungary.

King, M. (2006). *General principles of user oriented evaluation.* Kluwer Academic Publishers.

Kray, C., Elting, C., Laakso, K., & Coors, V. (2005). Presenting route instructions on mobile devices. In J. Dykes, A. M. Maceachren, & M. J. Kraak (Eds.), *Exploring geovisualization* (pp. 529-550). Elsevier, Amsterdam.

Langley, P., Laird, J. E., & Rogers, S. (2006). *Cognitive architectures: Research issues and challenges* (Tech. Rep.). Computational Learning Laboratory, CSLI, Stanford University, CA.

Lee, S. H. (1999). Usability testing for developing effective interactive multimedia software: Concepts, dimensions, and procedures. *Educational Technology & Society, 2*(2).

Lewis, J. R. (2001). Introduction: Current issues in usability evaluation. *International Journal of Human-Computer Interaction, 13*(4), 343-349.

Luo, L., & John, B. E. (2005). Predicting task execution time on handheld devices using the keystroke-level model**.** In *Proceedings of the International Conference on Human Factors in Computing System (CHI 2005*), Portland, OR.

Matessa, M., Vera, A., John, B., Remington, R., & Freed, M. (2002).Reusable templates in human performance modeling. In *Proceedings of the Twenty-fourth Annual Conference of the Cognitive Science Society.* Mahwah, NJ: Lawrence Erlbaum Associates.

Mayhew, D. (1999). *The usability engineering lifecycle*. San Francisco: Morgan Kaufman Publishers.

Newell, A. (1072). *Unified theories of cognition.* Harvard College Press.

Newell, A. (1973). You can't play 20 questions with nature and win: Projective comments on the papers of this symposium. In William G. Chase (Ed.), *Visual information processing* (pp. 283-308). New York: Academic Press.

Nielsen, J. (1994). Heuristic evaluation. In: J. Nielsen& R. L. Mack (Eds.), *Usability inspection methods.* New York: John Wiley & Sons.

Olson, G. M., & Moran, T. P. (1998). Introduction to this special issue of experimental comparisons of usability evaluation methods. *Human-Computer Interaction, 13*,199-201.

Oulasvirta, A., Tamminen, S, Roto, V., & Kuorelahti, J. (2005, April). *Interaction in 4-second bursts: The fragmented nature of attentional resources in mobile HCI.* Paper presented at Computer-Human Interaction, Portland, OR.

Peschl, M., & Stary, C. (1998). The role of cognitive modeling for user interface design representations: An epistemological analysis of knowledge engineering in the context of human-computer interaction. *Mind and Machines, 8,* 203-236.

Redmond-Pyle, D., & Moore, A. (1995). *Graphical user interface design and evaluation.* London: Prentice Hall.

Ritter, F. E., Baxter, G. D., Jones, G., & Young, R. (2000). Supporting cognitive models as users. *ACM Transactions on Computer-Human Interaction, 7*(2), 141-173.

Ritter, F. E., Klein, L. C., Quigley, K., & Schoelles, M. (2004). Using cognitive modeling to study behaviour moderators: Pre-task appraisal and anxiety. In *Proceedings of the Human Factors and Ergonomics Society* (pp. 2121-2125).

Ritter, F. E., Rooy, D., St. Amant, R., Simpson, K. (2006). Providing user models direct access to interfaces: An exploratory study of a simple interface with implications for HRI and HCI. In *IEEE Systems, Man, and Cybernetics: Part A, Systems and Humans (to appear).*

Ritter, F. E., & Young, R. M. (2001). Embodied models as simulated users: Introduction to this special issue on using cognitive models to improve interface design. *International Journal of Human-Computer Studies,* 1-14.

Rubin, J. (1994). *Handbook of usability testing. How to plan, design, and conduct effective tests.* New York: John Wiley & Sons, Inc.

Salvucci, D. D., & Macuga, K. L. (2001). Predicting the effects of cellular-phone dialing on driver performance [Special Issue].*Cognitive Systems Research*.

Schneiderman, B., & Plaisant, C. (2005). *Designing the user interface. Strategies for effective human-computer interaction* (4th ed.). Boston: Pearson.

St. Amant, R., & Riedl, M. O. (2000). A perception/action substrate for cognitive modeling in HCI. *International Journal of Human-Computer Studies, 55,* 15-39.

St. Amant, R., Horton, T. E., & Ritter, F. E. (2006). Model-based evaluation of expert cell phone menu interaction. *Proceedings of the ACM Conference on Human Factors in Computing Systems CHI '04*, Vienna .

Sun, R. (2004). Desiderata for cognitive architectures. *Philosophical Psychology, 17*(3), 341-373.

Taatgen, N.A., Lebiere, C., & Anderson, J. R. (2006). Modeling paradigms in ACT-R. In R. Sun (Ed.), *Cognition and multi-agent interaction: From cognitive modeling to social simulation* (pp. 29-52). Cambridge University Press.

Treu, S. (1994). *User interface evaluation. A structured approach*. New York: Plenum Press.

Zetie, C. (1995). *Practical user interface design. Making GUIs work*. New York: McGraw-Hill, Inc.

KEY TERMS

Cognitive Architecture: Theoretical framework containing a set of relatively independent core constraints that are relatively constant across tasks and time

Cognitive Model: Any external representation of mental states

Cognitive Modeling: Mapping of the mental models onto cognitive models

Computational Cognitive Models: Programmable cognitive models that run on a computer

Interactive Behaviour: Type of behaviour that emerges from *dynamic interaction* among the elements, such as those within a user-task-artefact triad

Usability: In a general ergonomic sense, it is a product's fitness for purpose

Usability (of a software product): A specific software quality characteristic, usually analyzed in terms of effectiveness, efficiency, and users' satisfaction with the product

Quality in Use: User's impression on a software product's quality

Chapter XLIX
Assessing Human Mobile Computing Performance by Fitts' Law

Thomas Alexander
FGAN - Research Institute for Communication, Information Processing, and Ergonomics, Germany

Christopher Schlick
RWTH Aachen University, Germany

Alexander Sievert
German Sport University Cologne, Germany

Dieter Leyk
German Sport University Cologne, Germany
Central Institute of the Federal Armed Forces Medical Services Koblenz, Germany

ABSTRACT

This chapter describes the interdependence between locomotion while walking and human input performance in mobile Human-Computer-Interaction (HCI). For the analysis of the interdependence, appropriate performance measures, for example, subjective workload ratings or error rate, have to be applied. The way in which Fitts' law can enhance the analysis is explained. In an experiment with n=18 participants, the general indices of performance (bits per second) were measured while standing and walking with constant speed (2, 3.5, 5 km/h). Results show a significant increase of the error rate and a significant decrease of the index of performance for increased walking speed. Subsequent regression analyses allow quantitative estimation of these effects. The results show a division of the interdependence in two parts, based on the difficulty of the input task; they define threshold values for accuracy of user input. These values can be applied for the implementation and design of future Graphical User Interfaces (GUI) for mobile devices.

INTRODUCTION

Flexibility, variability, and mobility are topics of growing importance for today's society. This trend affects work with modern IT-systems (Goth, 1999): There is a growing availability and market for portable and mobile devices. They facilitate ubiquitous information access throughout customers' visits, while traveling, wandering through a production plant, or for working at home offices. It is expected that the market share of telecommuting and according devices for information access will increase. IT-developers and providers share this optimistic estimation of the growth potential (Business Week, 2006). They assume that today's mobile computers already have a market share of 40% (Microsoft, 2006). This requires special information infrastructures and personal mobile devices.

Common portable and mobile devices are notebooks, tablet-PCs, personal digital assistants (PDA), and so-called smartphones, which are cellular phones with enhanced functionality. For applications while standing or on the move (walking), when no tables or horizontal racks are available, weight and size issues are most relevant. They reduce the available devices to small, lightweight PDAs and smartphones.

PDAs and especially smartphones often rely on direct keypad input. Keypads allow a fast selection of a limited number of special functions. However, this is hardly sufficient for a more complex interaction. In that case, a point-and-click procedure is applied, which requires special touch-sensitive screens and pens for HCI and a WIMP-metaphor (windows, icons, menus, and pointer) is implemented (well known from most desktop systems) for the graphical user interface (GUI). Required training is reduced and most users can instantly use the device. Text input is facilitated by a miniaturized (virtual) keyboard or handwriting recognition. The keyboard solution displays a miniaturized QWERTY-keyboard and keys are selected by pointing and clicking. Handwriting recognition requires a stable position of the base for a precise text input. In both cases, pointing and pointing accuracy are essential.

Most of today's GUIs of mobile devices are simply adapted from stationary desktop systems. Characteristics of mobile use and their effects have not been considered (Berteksen & Nielsen, 2000; Crowley et al., 2000; Danesh et al., 2001; Dunlop & Brewster, 2002; Lumsden & Brewster, 2003; York & Pendharkar, 2004). However, effects of mobility on input performance are likely because of various reasons. First, walking itself causes distracting movements and forces on the arm and hand system. This leads to reduced input performance. Second, pointing and moving are two concurrent tasks, both of which require attention and processing resources. As a matter of fact, performance in either of the two tasks is reduced. This can be observed when users either stop when working with a mobile device or quit working with the device.

These observations show an overall need for the inclusion of general ergonomic findings, results, and models to optimize HCI and the according GUI on the move. As a first step, valid measures have to be analyzed in order to quantify input performance under different mobile conditions. They must be sensitive enough to detect even small effects. Based on these measures, subsequent analyses will give more detailed recommendations for the design of the GUI. This way a real mobile use can be achieved.

ASSESSING INPUT PERFORMANCE OF MOBILE DEVICES

Using computers on the move is a combination of walking and HCI. There are several reasons to assume interdependences between both. The extent of these interdependences varies with the degree of mobility. For quantifying it is necessary to identify and measure HCI performance correctly. There are different methods for doing so. This section addresses these issues and proposes the inclusion of general characteristics of the human operator.

Mobility and HCI: Two Concurrent Tasks

Mobile computing consists of the parallel processing of two tasks: walking or being "on the move" on the one hand, and HCI on the other. Such parallel processing results in interference between the two tasks (Hinckley et al., 2000; Navon & Gopher, 1979; Wickens, 1984). The extent of the interference varies with the similarity of the task.

Walking is a complex task. It requires multimodal (visual, auditory) encoding of environmental stimuli, complex central processing for reacting to the encoded stimuli and navigational orientation, and finally, controlled and coordinated motor output (i.e., walking itself). Most people are very well trained in performing this complex task so that several sub-tasks are delegated to lower processing levels. Posture control is one of those. When walking becomes more difficult due to higher speed or obstacles, a higher level of attention is required. This leaves few or no resources for additional processing (Baber et al., 1999; Kristoffersen & Ljungberg, 1999; Oulasvirta, 2005).

Interacting with the computer also requires visual encoding, cognitive processing, and motor responding. The extent of resource exploitation varies with task difficulty. While simple tasks can be handled with little attention, difficult or more complex tasks require a lot of attention and processing resources.

Consequently, both tasks compete for similar attentional and processing resources. This hypothesis is supported by the following observation: If walking speed increases or walking becomes more difficult in an environment with a lot of obstacles, people quit working with the PDA. This is because walking demands the major part of the available resources. These resources are drawn from HCI, which is the secondary task. However, if HCI becomes the main task because of prioritization, people often stop walking. In this case, available resources are moved to HCI. Too few resources are left for the walking task.

In general, the shifting of resources towards one task or the other is a deliberate decision. It is influenced by various factors (motivation, deadlines, instruction, etc.), which are beyond conscious control. However, due to safety reasons and the importance of avoiding accidents, walking is usually the primary task and HCI the secondary. This is especially true for moving in traffic situations. The highly dynamic, stimuli-rich environment causes a reduction of the available resources. There are situational induced impairments and disabilities (SIID) which hinder mobile computing (Sears et al., 2003). During movements, spatial orientation and self-movements are clearly main tasks, and only limited attention and processing resources are available for interacting with the computer (Kristoffersen & Ljungberg, 1999; Lumsden & Brewster, 2003; Pascoe et al., 2000; Perry et al., 2001). There are many attentional shifts between environment and task processing on the mobile device (Oulasvirta, 2005).

In addition to resource sharing, there is a biomechanical coupling between walking and HCI. The walking movement itself causes additional interfering forces on the hand-arm-system. They are permanently present during walking. The result is a task that includes hitting a target on the moving display with a moving input device. This is much more difficult than stationary HCI and requires additional processing.

There are many further interrelationships between walking and mobile computing. Ebersbach (1995) found changes of gait kinematics and reduced task performance while walking. Targets on the display are more frequently missed while walking (Beck, 2002). In experiments, Brewster (2002) found lower workload and higher input performance for standing rather than walking. This increases with complexity of the task. Although there are no differences between stationary and mobile processing of single tasks, task performance of more complex tasks varies (Barnard et al., 2005).

Mobile Input Performance

Consequently, input performance for evaluating mobile computing should be examined in a realistic setting, ideally during walking. However, an extensive research of Kjeldskov & Stage (2004) reveals that this was only done in few publications, referring to an evaluation of a mobile system and an acoustic navigation system for blind persons. Our similar research identified a few more publications which found limited usability, reduced task performance, and higher workload for mobile computing (Barnard et al., 2005; Beck et al., 2002; Brewster, 2002; Lin et al., 2005; Oulasvirta, 2005).

In general, there are two different methods for analyzing mobile input performance: laboratory studies and field experiments. In a quick comparison, the latter obtain higher face validity because measures are taken under real-life conditions (Thomas et al., 2002). Field experiments allow real walking, where kinematics differ from those when walking on treadmills (Nigg et al., 1995; Schache et al., 2001; Vogt et al., 2002). With treadmills, step length is reduced and different hip, ankle, and spine motions occur (Arlton et al., 1998; Murray, 1985).

In contrast to this, laboratory tests offer a wider control of environmental conditions (Alton et al., 1998; Johnson, 1998; Kjeldskov et al., 2004; Pirhonen et al., 2002). Key situations can be analyzed relatively simply and some experimental methods such as observation or thinking aloud can be applied (Rantanen et al., 2002; Sawhney & Schmandt, 2000). The participant is in a safe state throughout the experiment. This cannot always be guaranteed in all environments in field experiments (Kjeldskov & Stage, 2004). To gain more general insight into mobile interaction performance, laboratory studies are selected rather than field experiments (Alton et al., 1998; Kjeldskov & Graham, 2003; Petrie et al., 1998).

There are many ways for measuring human input performance. A common method for assessing input performance is subjective rating. In this case, participants are simply interviewed after the experiment about their own performance (Barnard et al., 2005; Mustonen et al., 2004). The participants' comments are recorded and transferred into a qualitative or quantitative measure afterward. Questionnaires are another method for assessing subjective input performance. Here, questions about special qualities are presented to the participants. The participants estimate numerical values for the extent of this quality. There are standardized questionnaires available for measuring general usability, workload, and so forth (Barnard et al., 2005; Beck et al., 2002; Kjeldskov & Stage, 2004). A further method is observation by experimenters (Jacobsen et al., 2002; Kjeldskov & Stage, 2004; Oulasvirta, 2005). The participants are observed while performing their specific HCI task. Observers report characteristics, categorize, and rate characteristics. This method seems to be more objective than the ratings described before. However, observations remain subjective because they are strongly dependent on the observer.

For an objective assessment of input performance, various performance variables can be applied. Most of them are fully integrated into the main experimental task and refer to accuracy, time, and task-specific measures.

A common measure for accuracy is error rate (r_E), which is defined as the ratio of missed targets to total targets. Obviously, higher error rates mean reduced performance. Barnard et al. (2005) and Beck et al. (2002) report higher error rates for mobile than for stationary use. Other error measures are more task-dependent and refer to wrong command selection or wrong ways of task processing. It must be noted that error rate is always dependent on the difficulty of the task. If a task is simple, the error rate is low, while a difficult task results into a higher error rate. There is a close dependence between accuracy and time. Obviously, tasks are performed more accurately when more time is taken and vice versa. Consequently, accuracy measurements should involve an explicit prioritization in the experimental task and task times should be recorded as well.

In terms of time, the time to complete a special task or sub-task is frequently used (e.g., Barnard et al., 2005; Cerney et al., 2004; Kjeldskov & Stage, 2004). Time measurement starts with the task and ends as soon as the task is completed. Temporal performance measures are typically task-dependent and serve as intermediate results for more general findings. When recording temporal measures, it is important to consider and determine the overall temporal accuracy of the system. This is especially important for mobile devices. Although today's devices run with high-speed CPUs, the system update is not identical to the maximum resolution of temporal measures. The actual measurement rate is smaller because the systems are not real-time systems and there are many basic I/O system processes running in parallel. Moreover, at least the double measurement frequency is required for an exact representation of the measured signal (Nyquist, 1928).

The last category includes task-specific measures. Such measures are more complex and refer to complex problem-solving approaches (Dahm et al., 2004; Oppermann, 2003; Ziefle, 2002). An example for a task-specific measure is linguistic processing and semantic understanding. In this case, a special scenario is described on the mobile display and the participant answers questions about it. Evidently, time to complete the special task and error rate are dependent on the application. Results of these task-specific measures are often difficult to generalize.

Applying Fitts' Law for Assessing HCI Performance

For a more comprehensive quantitative analysis of mobile HCI performance, general laws of human information processing have to be referenced. A basic law that was first discovered by Fitts (1954) describes the relation between movement time and index of difficulty for goal-directed movements. Thus, it facilitates an estimation of human sensorimotor performance. Fitts' law has been frequently applied for researching HCI issues of many kinds. Applying it for mobile GUIs is a way of quantifying human performance under mobile conditions.

Common user input on a PDA includes pointing and clicking on targets. The necessary movements are goal-directed and visually controlled arm-hand movements. For estimating the performance of these movements, Fitts' law can be applied (Fitts, 1954; Fitts & Peterson, 1964). According to this, the movement time (MT) is linearly dependent on the index of difficulty (ID) of a movement. It is:

$$MT = a + b\ ID \qquad (1)$$

Fitts' coefficients a and b are determined by regression. The first coefficient is a theoretical intercept for ID=0. It can be interpreted as reaction time. The second coefficient b is a measure of input performance and it characterizes the slope of the curve. Its reciprocal value (1/b) describes the index of performance in bits per second (bps) (Card et al., 1978). The ID of a movement is the logarithm of the quotient of target width (W) and amplitude of the movement (A). It is:

$$ID = \log_2 (2\ A/W) \qquad (2)$$

Goal-directed movements are divided into an initial ballistic and a final visually controlled phase with different characteristics. The ballistic phase is fast and inexact. Its purpose is to move the finger towards the target region as fast as possible. The visually controlled phase takes longer but is more precise. It is used to hit small targets. The ID determines the more dominant phase of the movement. Simple movements with ID<3.58 are primarily ballistic, while more difficult movements with higher ID are visually-controlled (Wallace & Newell, 1983). Gan & Hoffmann (1988) found a threshold between ID=3 and 4 for the transition between ballistic and visually controlled movements.

Fitts' law has been used for many analyses of goal-directed movements, several of which have a background in HCI (compare MacKenzie, 1995), and allows a quantification of input performance for precise manual inputs.

However, despite the frequent applications of Fitts' law for stationary HCI, it has only rarely been used for investigating mobile HCI or different GUI layouts. In their publication, Lin et al. (2005) verify it for general pointing tasks on a PDA. There is a linear interrelationship between movement time and index of difficulty. It was determined that performance was slightly reduced with mobile conditions. The according indices of performance decrease from 7.8 bps (sitting), to 6.5 bps (slow walking), and, finally, to 6.6 bps (fast walking). In contrast to this, Cerney et al. (2004), with their study about different mobile text entry methods, found that Fitts' law was not applicable for PDA interaction. Measured movement times were independent from the index of difficulty. A reason for this was that the analyzed IDs were below 3. As stated before, this characterizes ballistic movements where the applicability of Fitts' law is limited. As of yet, there is no general consensus about the effect of the user's movement on Fitts' law and derived measurements of HCI performance.

Experiment on Mobile Input Performance on the Move

This work focuses on a common application of mobile devices: orientation and navigation in an unknown area. An electronic map of the surroundings is shown on the PDA-display. The user can select targets and the system calculates waypoints toward this target accordingly. A high level of precision is required for target input. For this application, user input performance should be analyzed while standing and during walking at different speeds. This ensures that the mobile characteristics are considered accordingly. The results, however, are not valid for this specific task alone. The resulting values are benchmarks for each point-and-click procedure for GUIs. In this case, they are used to determine size of icons and menu items for an optimized GUI.

Hypothesis

Walking and HCI share human information processing resources to a special extent. Higher processing demands on one task result in decreased performance in the other. If, for instance, walking becomes more demanding because of a higher walking speed, HCI performance will be reduced. It is postulated:

Walking affects input performance of a targeting task on a PDA. The effect varies with the extent of walking difficulty (i.e., walking speed).

Input performance can be characterized in general by two variables. It is accuracy/error rate r_E, on the one hand, and movement time on the other. Fast input leads to a reduced precision with a higher error rate and vice versa. For quantifying input performance it is necessary to consider both variables. In order to avoid misunderstandings between the different types of movements (walking and hand movement), the term input time (IT) will be used in the future, instead of movement time.

As a result of different IDs, there is a separation and transition between simple and difficult movements. In this study, this separation had to be considered for the analysis of r_E and IT. The separation might result in changes of the dependency between ID and input performance. If the results were to show such differences, the further analyses would be performed separately for each movement type.

For specifying the effect on error rate r_E, a general linear model is selected. The model with intercept c_E, and the coefficients c_v, c_{ID} and c_{vID} includes linear terms for walking speed v, for index of difficulty ID, and for the interaction of both terms v x ID. It is:

$$r_E(v, ID) = c_E + c_v\, v + c_{ID}\, ID + c_{vID}\, v\, ID \qquad (3)$$

Error rate and walking speed can be measured instantly.

For assessing human input performance, the according index of performance (1/b) is determined in [bps]. This requires additional pre-analyses. At first, separate regression analyses are performed in order to estimate the parameters of Fitts' law separately for each walking speed. A subsequent inclusive regression analysis considers additional terms for the influence of walking speed. Thus, both coefficients of Fitts' law, a and b, are split into a velocity-independent (index ID) and a velocity-dependent fraction (index v). It is:

$$IT(v, ID) = (a_{ID} + a_v\, v) + (b_{ID} + b_v\, v)\, ID \qquad (4)$$

Notice that the effect of ID is considered in the function already so that there is no need to include an additional term for the interaction of walking speed and ID. By comparing both linear coefficients b_{ID} and b_v, it is possible to put the effect of the velocity-dependent term and the velocity-independent term in perspective.

Participants

Eighteen members of the participating institutions took part in the experiments. Mean age was 29 +/- 10.6 yrs (AM +/- 1 sd). Fifteen participants were male, 3 female. Each participant was familiar with computer work, computer interaction, and interaction with a PDA. None was involved in designing this study or in preparing the specific experiment. Each participant completed the total experimental trials in a randomized, balanced order.

Apparatus

A standard Dell Axim X50v-Pocket PC was utilized for the experiment (Processor Intel Xscale PXA270, 624 MHz, dimensions: 73 x 119 x 18 mm, weight: 165 g, display diagonal: 9.4 cm, resolution: 480 x 640 pixel). Participants used the standard input pen (10 cm length, 3 mm diameter).

Figure 1. Participant and HP Cosmos Pulsar™ treadmill

A H/P/Cosmos Pulsar™ treadmill, as shown in Figure 1, was used for providing and controlling a constant walking speed. The size of the treading surface was 190 cm x 65 cm. For safety reasons, it was equipped with a safety belt, fall stop, and breast clamp. The device is certified for medical treatment.

Procedure

The experiment includes walking on the treadmill and a simultaneous HCI task. Each experiment starts with a HCI trial during standing. This trial serves as a baseline. Subsequent trials on the treadmill follow in a balanced randomized order. Walking speed is varied between 0 (i.e., standing), 2, 3.5, and 5 km/h (first experimental factor). Higher walking speeds are not considered because of safety reasons.

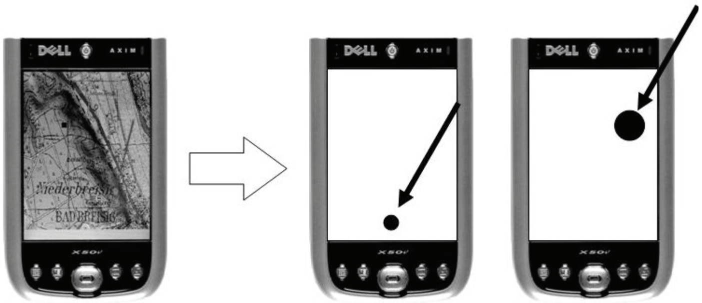
The information input task on the PDA is a target assignment task, as it is shown in Figure 2. It simulates interaction with navigation software on a PDA. During the experimental task, a starting point and a single target point appear sequentially. Participants are instructed to click the appropriate points. The size and position of the starting point remain constant. The size and position of the target vary.

Based on the results from pre-tests, circular target sizes from d=10 pixel (2.4 cm) to d=50 pixel (12 mm) are used. The distance between the starting and the target point is varied, so that according IDs vary from 2.0 to 5.6. The step size between each ID condition is 0.2 ID. This results in 19 steps for the ID (second experimental factor).

Participants are instructed to hit the target point as accurately as possible (prioritizing of precision against input speed). Each ID-step is repeated three times. In total, 57 single start-target movements are carried out per trial. The order of movements is randomized. After each trial the actual error rate [percent] is fed back to the participants. Three trials are carried out directly following each other in one session.

During the trials, relative error rate (missing target point) and time between appearance of the target point and hitting the touch-screen with the

Figure 2. Dell Axim™ PDA showing map (left) and experimental pointing task (middle: start position; right: target position)



pen in [ms] are measured. The accuracy of the time measurement is set to 55 ms.

A single session takes about 15 minutes. At the end of each session a break of at least 5 minutes follows. During the break, the participants rate the subjective workload or task difficulty on a two-level judgment scale (Käppler, 1993). The scale has verbal and numerical descriptors, which allows a differentiation between "simple" (value 0) and "difficult" (value 100) tasks.

Statistical Analysis

The statistical analysis was performed with the statistical software package SYSTAT™ Version 11.0 (Systat, 2004).

The error rate was analyzed by a two-way analysis of variance (ANOVA) with the two independent variables walking speed and ID. A subsequent pairwise comparison with Sidak adjustment was used to identify significant differences between the two different factor levels. When a separation into subgroup was revealed, the subsequent analyses were performed separately for each subgroup. The postulated relationship between relative error rate, ID, and walking speed was specified by a final multiple regression analysis for each of the subgroups.

The second part of the analysis referred to Fitts' law. At first, separated regression analyses tested the applicability and validity of Fitts' law for each walking speed. Errors were omitted from the analyses, as proposed by Card et al. (1978). Afterwards, the individual function for input time and index of performance (1/b) were calculated for each participant. Differences were tested for significance by a one-way ANOVA (factor: walking speed) and subsequent pairwise comparison (including Sidak adjustment). In case a grouping between factor levels appeared, the subsequent analyses were performed separately for each group. For the specification of the relationship between input time, ID, and walking speed, a final multiple regression analysis was carried out.

The subjective ratings were tested for significance by a one-way ANOVA (factor: walking speed) and subsequent pairwise comparisons (including Sidak adjustment). The results were used to compare the sensitivity of subjective ratings to the sensitivity of the objective ratings mentioned.

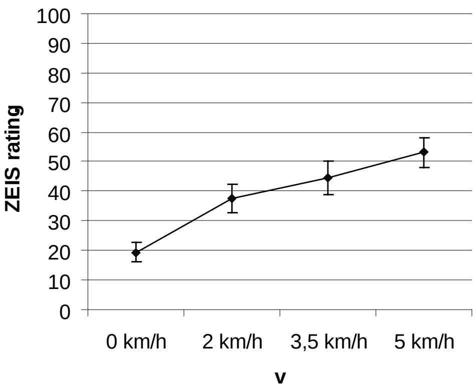
The chosen level of significance for each analysis was p=0.05.

RESULTS

Subjective Workload

The ratings for the subjective workload and task difficulty increase from 19 (standing) to 53 (5 km/h). This corresponds to an increase from "easy" to "predominantly difficult." Workload differs between standing and walking. There is only a small linear increase with increased walking speeds. In this case, the ratings are similar. The according distributions are shown in Figure 3.

Figure 3. Mean and standard error of subjective workload (ZEIS rating) at four walking speeds (v)



The ANOVA reveals a significant effect of walking speed on the subjective rating ($F_{3,51}$=13.5; p<0.01). The factor explains o^2=41% of the observed variance.

A final pairwise comparison put this result into perspective. It shows only significant differences between standing and walking (p<0.01), but not between different walking speeds. These results show that the amount of subjective workload is similar during walking. It does not increase for walking speeds between 2 km/h and 5 km/h.

Error Rate

Error rate was measured for each walking speed and each of the different IDs. Figure 4 shows the distribution of the relative error rates for each walking speed level.

For simple interaction tasks, no errors occurred. Therefore, two conditions were merged into a single group for the subsequent ANOVA. Both factors, ID ($F_{9,153}$=36.0; p<0.01) and walking speed ($F_{3,51}$=15.7; p<0.01) have a significant effect on error rate. It increases from 4.2% (standing) to 10.1% (5 km/h) on the average. The interaction of both factors affects error rate as well ($F_{27,459}$=2.0; p<0.01). The factor ID explains o^2=80.6% of the variance, while the factor walking speed only explains o^2=9.3%. The interaction explains o^2=4% of the variance.

However, the subsequent pairwise comparison reveals significant differences only between two different ID-sections. There are no significant differences of error rate within each section. One section includes simple movements with ID<4 and the other, more difficult movements with ID≥4. This result makes a separate analysis for each section necessary.

For movements with ID<4 the relative error rate is nearly constant. It varies between 0% and 4.3% for all IDs and walking speeds. As shown in figure 3, there is a linear effect of both factors. Both factors, ID ($F_{4,68}$=3.7; p<0.01) and walking speed ($F_{3,51}$; p<0.01) affect error rate. There is no significant interaction ($F_{12,204}$=0.5; p=0.9). Therefore, it will be omitted for the further analysis. The ID explains o^2=31.4%, and the walking speed o^2=34.8% of the variance.

The regression analysis estimates the relative error rate r_E based on ID and walking speed v. For movements with ID<4 it is r_E:

$$r_E(v, ID) = - 0.027 + 0.012\ ID + 0.003\ v \qquad (5)$$

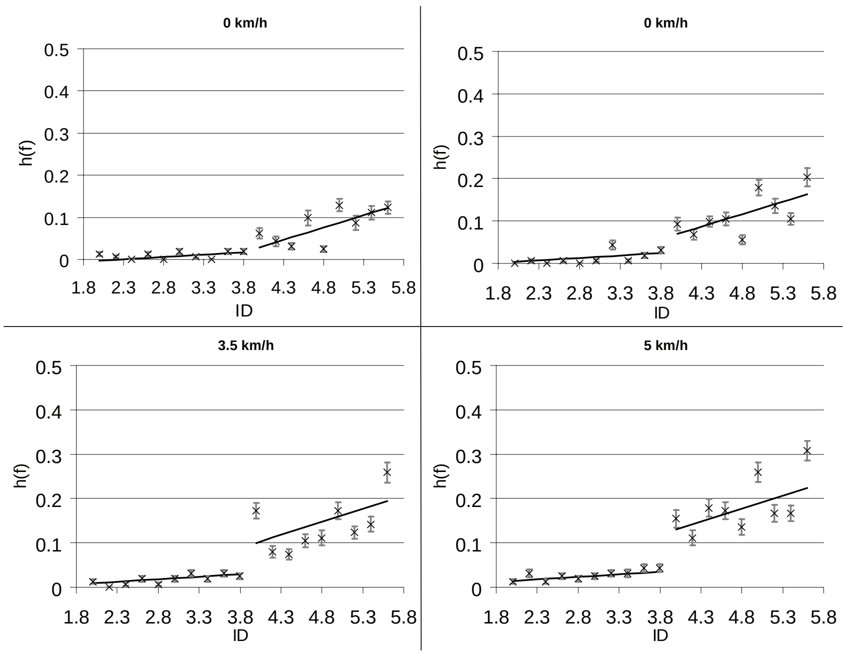
According to this regression function, error rates between 0% and 3.6% are expected for ID<4 and v<5 km/h. This is nearly a constant relationship.

The constant behavior of the error rate changes for movements with ID≥4. As shown in Figure 4, error rates of these movements increase much stronger with growing ID and walking speed. Between ID=4 and ID=5.6, the error rate doubles or increases even more. The ANOVA reveals a significant effect of ID ($F_{4,64}$=14.5; p<0.01) and walking speed (F3,51=15.5; p<0.01) on error rate. The according values of the explained variance are o^2=46.6% for ID and o^2=39.6% for walking speed. Since the interaction is not significant ($F_{12,204}$=1.2; p=0.30) it is omitted from the further analysis.

The regression analysis for estimating the relative error rate r_E for movements with ID≥4 determines the parameters of the model function as:

$$r_E(v, ID) = -0.208 + 0.059\ ID + 0.020\ v \qquad (6)$$

Figure 4. Mean, standard error, and regression function (separately for ID<4 and ID≥4) of error rate r_E for four walking speeds



This estimates absolute values for the analyzed inputs between 2.8% (ID=4; standing) and 22.2% (ID=5.6; 5 km/h). In comparison to movements with smaller IDs, the effect of walking speed is much stronger. This is because the coefficient of the walking speed increases with a factor of (0.02: 0.003 = 6.7) compared to an increase of (0.059: 0.012 = 4.9) of the coefficient of the ID.

Input Time and Fitts' Law

The initial correlation analysis for each walking speed condition reveals a high probability of a linear relationship between input time IT and ID ($r>0.955$; $r_{0.1\%}=0{,}693$). This confirms applicability of Fitts' law for the subsequent analysis.

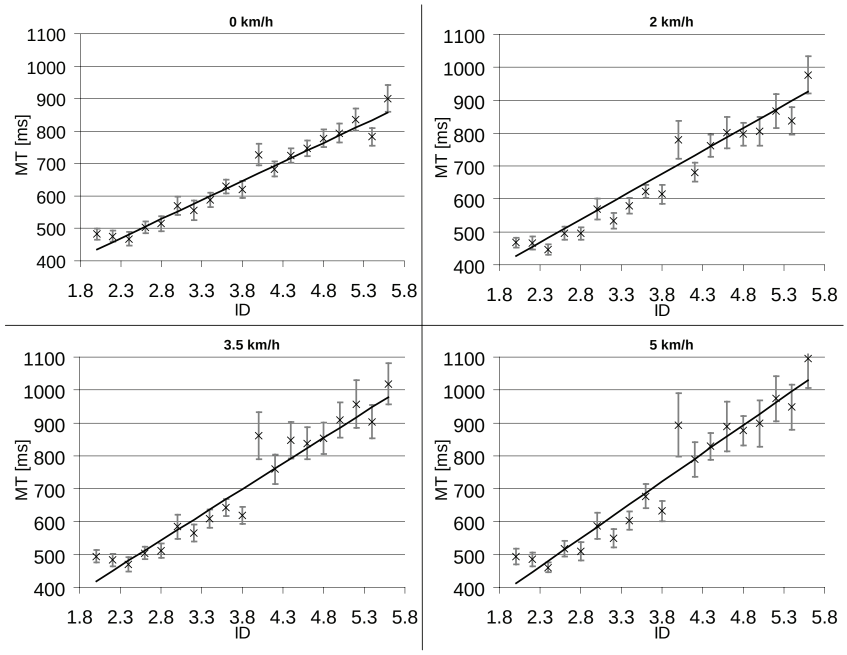
Figure 5 shows the empirical values for each walking speed. Notice that there are constant input times for interaction tasks with ID < 2.6. In contrast to the general linear relationship, no or just a small effect of ID or walking speed on movement time can be observed. For higher IDs the linear relationship is valid again. Apart from this observation, no evidence for a separation into two or more groups is found, so the following analysis refers to the total continuous data.

Based on the estimated Fitts' linear coefficient b, the index of performance (1/b) can be easily determined. It varies between 8.8 bps (standing) and 6.0 bps (5 km/h). Since the goal of this study was a specification of Fitts' law, the further analysis refers to the linear coefficient b again.

The ANOVA shows a significant effect of walking speed on the linear coefficient b ($F_{3,51}=6.7$; $p<0.01$). It explains $o^2=23.9\%$ of the observed variance. The following pairwise comparison reveals significant differences only for differences of walking speeds above 3 km/h. This can be observed throughout the total ID range. There are no evidences for differences that would require separate analyses.

The final regression analysis extends Fitts' law by additional terms, which are velocity-dependent. They specify the effect of the walking speed on movement time. Accordingly, input time IT in [ms] can be estimated by walking speed v and ID as:

Figure 5. Mean, standard error, and regression function of input time [ms] according to Fitts' Law for four walking speeds



$$IT\ (v, ID) = (198.4 - 26.2\ v) + (117.8 + 10.9\ v)\ ID \quad (7)$$

The according index of performance (1/b) for information input while standing is 8.5 bps. It decreases to 5.8 bps at a walking speed of 5 km/h.

The model function allows a comparison of the absolute effect between walking speed and ID. The effect of ID on input time is slightly stronger than the effect of walking speed. For a higher walking speed (v = 5 km/h) the according linear coefficient (b_2 = 10.9 x 5 = 54.5) reaches half the value of the velocity-independent coefficient (117.8). Although ID remains the dominant term, walking speed still has a large effect on input time. For high walking speed, input time increases to 1.5 times the initial input time when standing. This must be considered when designing HCI with more complex and difficult user input.

DISCUSSION

The analysis supports the use of objective variables to analyze the effect of motion status and walking speed on input performance. It is possible to quantify the effect of walking speed on mobile HCI performance by relative error rate and by applying Fitts' law for visually controlled movements.

In general, error rate increases with increased walking speed, as already observed by Barnard et al. (2005) and Beck et al. (2002). The experiment described in the previous section confirms the relationship and revealed a division of this relationship in two parts. For simple interaction tasks (ID<4), error rate remains practically constant and is affected little by ID and walking speed. The results show that both factors have a similar effect. However, error rate is lower than 4% throughout the range of both experimental factors.

The behavior for interaction tasks with ID≥4 is different. For these tasks, both factors have a stronger effect. This results in an increased error rate with up to 22%. The results show that walking speed has an effect similar to that of ID. This finding is important for practical GUI design since participants were explicitly instructed to focus on accuracy. In real applications, this is not likely. Consequently, even higher error rates can be expected.

A possible explanation for this separation into two parts is based on the division of movements into a ballistic and a visually controlled part. According to this, the main characteristic of a movement changes between ID=3 and ID=4 (Wallace & Newell, 1983; Gan & Hoffmann, 1988). Lower IDs are characteristically ballistic movements; higher IDs are visually controlled. A visually controlled movement requires more perceptual resources than a ballistic movement. Therefore, the parallel processing of a walking task leads to reduced PDA-input performance and higher error rates.

With regard to Fitts' law, there is no statistically significant division between the two parts, although there are some hints from the observation. Instead, a single, continuous function can be applied. This applicability of Fitts' law for describing interaction performance is similar to the results of Lin et al. (2005) with a simulated PDA, though it is in contrast to the findings of Cerney et al. (2004), who have not found a linear relationship between movement time and ID. An explanation for this might be Cerney's simple targeting tasks with ID<3. As shown in the previous section, there was also a constant error rate observed for ID<2.6. The constant behavior did not continue for more difficult movements with higher ID, though.

For the total experimental range of IDs between 2.0 and 5.6, Fitts' law was found to be applicable. Indices of performance could be determined based on the specification and the calculation of the constant and linear coefficient. The value of 8.5 bps for the standing condition is comparable to the value of 10.4 bps for computer mouse interaction (Card et al., 1978). It is higher than the values given in other studies for various interaction devices (compare MacKenzie, 1995). A comparison of the values with the empirical values of Lin et al. (2005) for a simulated PDA shows close similarity as well. In this study, indices of performance between 8.8 bps (standing) and 6.0 bps (5 km/h) were determined. Lin et al. (2005) observed indices of 7.8 bps (sitting), 6.5 bps (slow walking), and 6.6 bps (fast walking).

Referring to the subjective ratings, the participants were able to differentiate between "easy" (standing) and "predominantly difficult" (walking). Barnard et al. (2005) and Kjeldskov & Stage (2004) made a similar observation. Mustonen et al. (2004) found subjective ratings to be more sensitive than objective performance measures. In this study it was found that subjective ratings only allow a differentiation between motion statuses, but not between different walking speeds. There are two possible explanations for this. First, it could be that subjective ratings are not sensitive enough to detect changes. Second, subjective workload does not change between different walking speeds. However, both explanations postulate that the changes of workload caused by walking speed are relatively small.

The results support the initial hypothesis of an influence of walking speed on input performance. It is also shown that this effect is separated into two, possibly three sections with different characteristics. In the first section (ID<2.6), error rate and movement speed are nearly constant. They are independent from walking speed and ID. With increased ID (2.6≤ID<4), Fitt's index of performance decreases with higher walking speeds. Error rate remains nearly constant. With increasing visually controlled movements (ID≥4), error rate increases strongly.

CONCLUSION

Using mobile IT-devices on the move, so-called mobile computing, is a challenging topic which requires novel approaches for assessing HCI performance and optimizing GUI design. Simple performance measures such as subjective ratings, error rate, or task-dependent time measures can be used for rough qualitative estimations, but they are often not sufficient for quantitative analyses of greater detail.

It was found that Fitts' law can be successfully applied for this. It is possible to estimate input performance and input time for easy and difficult movements and to differentiate even between small effects. Fitts' law facilitates modeling input performance (e.g., input time) based on target size and movement distance. In contrast to this, simple performance measures such as error rate, are less sensitive. They just change for more difficult input movements with ID>4. Subjective ratings for workload allow only a rough distinction between standing and moving. These results argue in favor of applying Fitts' law and derived objective measures as sensitive variables for mobile GUI design.

A simple way to optimize HCI performance is to stop and stand. Attention and information processing resources shift accordingly and enhance performance. This can frequently be observed in reality when people stop walking as soon as a task gets more complicated and they have to shift information processing resources towards HCI. However, this behavior severely constrains working on the move.

From a designer's perspective, the GUI of a device has to fulfill two important requirements for mobile computing. First, the target areas (icons, menus) must be large enough so that the targeting movement is easy. It is recommended that the index of difficulty of the required movements should remain below ID=4. As a result, high error rates can be avoided and user movements affect general input performance only marginally. Second, an error-tolerant behavior of the software is required because of the increased error rates with increasing walking speed. In the present experiment, error rates reached up to 22%, which hampers practical usability. It must be considered that input accuracy is usually prioritized in laboratory experiments. In reality, environmental stimuli are likely to significantly disturb the user.

The threshold value for ID defines the minimum size of target areas as icons or menus items. For pointing and selecting tasks it can be used to define "snap" areas. In contrast to current GUIs for PDAs which require a precise hitting of the target, the closest target within the snap area of the pen position is assigned.

This shows that a simple adaptation of desktop paradigms as GUIs must be considered insufficient for mobile computing. Moreover, it is necessary to develop and specify new paradigms which take the mobile environment and its special characteristics into consideration. Otherwise, high input performance requires stopping and standing. This inhibits real mobility and strongly contradicts the need for mobility in our society. General relationships from information processing can be applied as quantitative measures for defining prospective ergonomic design criteria and for evaluating new GUIs. Thereby, effective and efficient mobile computing on the move is facilitated.

FUTURE TRENDS

As described in the introduction of this chapter, mobility and the support of mobile working will gain importance in future. It will affect mobile computers and put high demands on mobile HCI. GUIs are required that take the characteristics of both mobility and the user's self-movement, into account. This results in the consideration of many aspects, ranging from the devices' weights and the dynamics of self-movement (walking) to sharing information processing resources. A great deal of research is still required in order to optimize GUIs for maximum human performance. There is a need for such studies since mobile devices such as PDAs or smartphones are broadly available and already used.

Furthermore, valid and accurate measures of human performance are required for prospective ergonomic GUI design and retrospective evaluation of implementations. This is not limited to specific measures, but refers to the general experimentation method, whether it is in laboratory studies or field experiments. The experiment described in this chapter was performed in a laboratory setting with a treadmill. Walking speed was controllable and additional environmental stimuli were reduced to a minimum. In contrast to such pre-defined settings, individual walking speeds will have to be analyzed in field experiments in the future. Since the method of applying measures such as subjective workload, error rate, and input time has been found to be successful and sensitive, a similar method may be applicable for outdoor and field experiments. By comparing the respective performance measures it will be possible to estimate and quantify the effect of the experimentation method (treadmill vs. field experiment).

More research is needed to specify similar requirements for HCI and GUI design. The experiment described in this chapter can be considered as an example of such a study. It gives detailed recommendations for the size of target areas that affect icon and menu item size. Another relevant aspect is visualization, because GUIs serve to visualize information as well. Daily experience shows that visual perception and cognitive processing is not independent from other environmental stimuli. Movement is likely to affect visual perception performance as well, and further empirical analyses are needed to specify the interdependence. Thereby, it will be possible, for instance, to define minimum icon and font sizes for mobile computing, independently from mobility status and walking speed. This is a more elegant way than just maximizing sizes or having the users adjust them manually (which is practically limited to changing font size).

In general, mobile HCI-analyses have to take characteristics of portability, that is, computing at different locations, and mobility, that is, computing while on the move, into account. By applying objective measures of performance, such as error rate or Fitts' law, it is possible to define ergonomic requirements for GUI design. Moreover, it is possible to evaluate different designs and optimize them.

REFERENCES

Alton, F., Baldey, L., Caplan, S., & Morrisey, M. C. (1998). A kinematic comparison of overground and treadmill walking. *Clinical Biomechanics, 13*(6), 434-440.

Baber, C., Knight, J., Haniff, D., & Cooper, L. (1999). Ergonomics of wearable computers. *Mobile Networks and Applications, 4*(1), 15–21.

Barnard, L., Yi, J. S., Jacko, J. A., & Sears, A. (2005). An empirical comparison of use-in-motion evaluation scenarios for mobile computing devices. *International Journal of Human-Computer Studies, 62*, 487-520.

Beck, E. T., Christiansen, M. K., & Kolbe N. (2002). *Metoder til Brugbarhedstest af Mobile Apparaten*. Aalborg University: Department of Computer Science (in Danish).

Bertelsen, O. W., & Nielsen, V. (2000). Augmented reality as a design tool for mobile interfaces. In *Proceedings of the Conference on Designing Interactive Systems: Processes, Practices, Methods, and Techniques* (pp. 185–192).

Brewster, S. A. (2002). Overcoming the lack of screen space on mobile computers. *Personal and Ubiquitous Computing, 6*(3), 188–205.

Business Week. (2006, March 2). *Why Google's going mobile*.

Card, S. K., English, W. K., & Burr, B. J. (1978). Evaluation of mouse, rate-controlled isometric joystick, step keys, and text keys for text selection on a CRT. *Ergonomics, 21*, 601-613.

Cerney, M. M., Mila, B. D., & Hill, L. C. (2004). Comparison of mobile text entry methods. In *Proceedings of the Human Factors and Ergonomics Society 48th Annual Meeting.* Santa Ana: HFES.

Crowley, J. L., Coutaz, J., & Blerard, F. (2000). Perceptual user interfaces: things that see. *Communications of the ACM, 43*(3), 54–64.

Dahm, M., Felken, C., Klein-Bösing, M., Rompel, G., & Stroick, R. (2004). Handyergo: Breite Untersuchung über die Gebrauchstauglichkeit von Handys. In R. Keil-Slawik, H. Selke, & G. Szwillus (Eds.), *Mensch & Computer 2004: Allgegenwärtige Interaktion* (pp. 75–84). München: Oldenbourg Verlag (in German).

Danesh, A, Inkpen, K., Lau, F., Shu, K., & Booth, K. (2001). Geney: Designing a collaborative activity for the palm handheld computer. In *Proceedings of CHI 2001* (pp. 388-395). New York: ACM.

Dunlop, M., & Brewster, S. (2002). The challenge of mobile devices for human-computer interaction. *Personal and Ubiquitous Computing, 6*, 235-236.

Ebersbach, G., Dimitrijevic, M. R., & Poewe, W. (1995). Influence of concurrent tasks on gait: A dual task approach. *Perceptual and Motor Skills, 81*, 107-113.

Fitts, P. M. (1954). The information capacity of the human motor system in controlling the amplitude of movement. *Journal of Experimental Psychology, 47*, 381-391.

Fitts, P. M., & Peterson, J. R. (1964). Information capacity of discrete motor responses. *Journal of Experimental Psychology*, *67*, 103-112.

Gan, K. C., & Hoffman, E. R. (1988). Geometrical conditions for ballistic and visually controlled movements. *Ergonomics*, *31*, 829-839.

Goth, G. (1999). Mobile devices present integration challenges. *IT Pro*, 11–15.

Hinckley, K., Pierce, J., Sinclair, M., & Horvitz, E. (2000). Sensing techniques for mobile interaction. In *Proceedings of the 13th Annual ACM Symposium on User Interface Software and Technology* (Vol. *2*(2), 91–100.

Initiative D21. (2006). *Press release of the D21-innovation workshop on mobile society*. Retrieved January 26, 2006, from http://www.initiatived21.de

Jacobsen, C. S., Langvad, P., Sorensen, J. J., & Thomsen H. B. (2002). *Udvikling og Vurdering af Brugbarhedstest til Mobile Apparater*. Aalborg University: Department of Computer Science (in Danish).

Johnson, P. (1998). Usability and mobility: interactions on the move. In *Proceedings of the First Workshop on Human-Computer Interaction with Mobile Devices*.

Käppler, W. D. (1993). *Beitrag zur Vorhersage von Einschätzungen des Fahrerverhaltens*. Fortschritt Berichte VDI-Reihe 12, Nr. 198. Düsseldorf: VDI-Verlag (in German).

Kjeldskov, J., & Graham, C. (2003). A review of mobile HCI research methods. In *Proceedings of the 5th International Mobile HCI 2003 conference*, Udine, Italy. Lecture notes in computer science. Berlin: Springer.

Kjeldskov, J., Skov, M. B., Als, B. S., & Hoegh, R. T. (2004). Is it worth the hassle? Exploring the added value of evaluating the usability of context-aware mobile systems in the field. In *Proceedings of the 6th International Mobile HCI Conference*, Glasgow.

Kjeldskov, J., & Stage, J. (2004). New techniques for usability evaluation of mobile systems. *International Journal Human-Computer Studies*, 599-620.

Kristoffersen, S., & Ljungberg, F. (1999). Making place to make IT Work: empirical explorations of HCI for mobile CSCW. In *Proceedings of the International Conference on Supporting Group Work (GROUP '99)* (pp. 276–285).

Lin, M., Price, K. J., Goldman, R., Sears, A., & Jacko, J. (2005). Tapping on the move—Fitts' law under mobile conditions. In *Proceedings of Information Resources Management Association International Conference* (pp. 132-135). San Diego, CA.

Lumsden, J., & Brewster, S. (2003). A paradigm shift: alternative interaction techniques for use with mobile & wearable devices. In *Proceedings of the 2003 Conference of the Centre for Advanced Studies on Collaborative research (CASCON'03)* (pp. 197-210). Toronto, Canada.

MacKenzie, I. S. (1995). Movement time prediction in human-computer interfaces. In R. M. Baecker, W. A. S. Buxton, J. Grudin, & S. Greenberg (Eds.), *Readings in human-computer interaction* (2nd ed.) (pp. 483-493). Los Altos, CA: Kaufmann.

Microsoft. (2005). Press release: *Go mobile*. Retrieved March 05, 2006, from http://msdn.microsoft.com/windowsvista/mobile/

Murray, M. P., Spurr, G. B., Sepic, S. B., Gardner, G. M., & Mollinger, L. A. (1985). Treadmill vs. floor walking: Kinematics, electromyogram, and heart rate. *Journal of Applied Physiology, 59*, 87-91.

Mustonen, T., Olkkonen, M., & Hakkinen, J. (2004). Examining mobile phone text legibility while walking. In *Proceedings of CHI 2004* (pp. 1243-1246). Vienna: ACM Press.

Navon, D., & Gopher, D. (1979). On the exonomy of the human processing system. *Psychological Review, 86*, 214-255.

Nigg, N. M., DeBoer, R., & Fisher, V. (1995). A kinematic comparison of overground and treadmill running. *Medicine and Science in Sports and Exercise, 27*, 98-105.

Nyquist, H. (1928). Certain topics in telegraph transmission theory. *Trans. AIEE, 47*, 617-644.

O'Donnell & Eggemeier. (1986). Workload assessment methodology. In K. R. Boff, L. Kaufman, & J. Thomas (Eds.), *Handbook of perception and human performance: Cognitive processes and performance*, vol. 2. New York: Wiley.

Oppermann, R. (2003). Ein Nomadischer Museumsführer aus Sicht der Benutzer. In G. Szwillus, J. Ziegler (Eds.), *Mensch & Computer 2003: Interaktion in Bewegung* (pp. 31-42). Stuttgart: B. G. Teubner (in German).

Oulasvirta, A. (2005). The fragmentation of attention in mobile interaction, and what to do with it. *Interaction, 12*(6), 16-18.

Pascoe, J., Ryan, N., & Morse, D. (2000). Using while moving: HCI issues in fieldwork environments. *ACM Transactions on Computer–Human Interaction (TOCHI), 7*(3), 417–437.

Perry, M., O'Hara, K., Sellen, A., Brown, B., & Harper, R. (2001). Dealing with mobility: understanding access anytime, anywhere. *ACM Transactions on Computing–Human Interaction (TOCHI), 8*(4), 323–347.

Petrie, H., Johnson, V., Furner, S., & Strothotte, T. (1998). Design lifecycles and wearable computers for users with disabilities. In *Proceedings of the First Workshop on Human-Computer Interaction with Mobile Devices*, GIST Technical Report G98-1, Glasgow.

Pirhonen, A., Brewster, S. A., & Holguin, C. (2002). Gestural and audio metaphors as a means of control for mobile devices. In *Proceedings of CHI' 2002*. New York: ACM.

Rantanen, J., Impio, J., Karinsalo, T., Reho, A., Tasanen, M., & Vanhala, J. (2002). Smart clothing prototype for the Artic environment. *Personal and Ubiquitous Computing, 6*, 3-16.

Sawhney, N., & Schmandt, C. (2000). Nomadic radio: speech and audio interaction for contextual messaging in normadic environments. *Transactions on Computer-Human Interaction, 7*(3), 353- 383.

Schache, A. G., Blanch, P. D., Rath, D. A., Wrigley, T. V., Starr, R., & Bennell, K. L. (2001). A comparison of overground and treadmill running for measuring the three-dimensional kinematics of the lumbo-pelvic-hip complex. *Clinical Biomechanics, 16*(8), 667-680.

Sears, A., Lin, M., Jacko, J., & Xias, Y. (2003). When computers fade … pervasive computing and situationally induced impairments and disabilities. In C. Stephanidis & J. Jacko (Eds.), *Human-computer interaction: Theory and practice* (Part II) (pp. 1298-1302). London: Lawrence Erlbaum Associates.

SYSTAT. (2004). *SYSTAT Version 11.* SYSTAT Software Inc. Richmond, CA.

Thomas, B., Grimmer, K., Zucco, J., & Milanese, S. (2002). Where does the mouse go? An investigation into the placement of a body-attached TouchPad mouse for wearable computers. *Personal and Ubiquitous Computing, 6*, 97-112.

Vogt, L., Pfeifer, K., & Banzer, W. (2002). Comparison of angular lumbar spine and pelvis kinematics during treadmill and overground locomotion. *Clinical Biomechanics, 17*(2), 162-165.

Wallace, S. A., & Newell, K. M. (1983). Visual control of discrete aiming movements. *Quarterly Journal of Experimental Psychology*, *35a*, 311-321.

Wickens, C. D. (1984). Processing resources in attention. In R. Parasuraman & R. Davies (Eds.), *Varieties of Attention* (pp. 483-493). New York: Academic Press.

York, J., & Pendharkar, P. C. (2004). Human-computer interaction issues for mobile computing in a variable work context. *International Journal of Human-Computer Studies*, *60*, 771-797.

Ziefle, M. (2002). The influence of user expertise and phone complexity on perfomance, ease of use and learnability. *Behaviour & Information Technology*, *21*(5), 303–311.

KEY TERMS

Fitts' Law: Based on Shannon's theorem for information processing, *Fitts' law* allows the estimation of required times for rapid movements between a starting point and a target area. Fitts introduced the index of difficulty (ID) as a characteristic measure for such movements. The ID is defined as:

$ID = \log_2 (2 A / W)$.

The term A describes the amplitude of the movement, or the distance between starting point, and target area, and the term W, the width of the target area. The movement time MT is linearly dependent on the ID. It is:

$MT = a + b\ ID$.

The coefficients *a* and *b* are both regression coefficients. The coefficient [a] defines the intercept for ID=0 and the coefficient b the steepness of the relationship.

Index of Performance: The *index of performance* is a measure for characterizing the speed of a visually controlled movement. Thus, it is also a measure of movement performance. It is calculated by the reciprocal value of the linear coefficient b of Fitts' law. Another definition is the ratio of $ID_{average}$ to $MT_{average}$. Both definitions allow a comparison of input performance under different movement conditions.

Mobile Computing: In contrast to stationary computing, which comprises stationary working with a computer at the same location, and portable computing, which refers to stationary working with a computer at different locations, the term *mobile computing* describes working with a computer while moving. This leads to weight and size requirements and special demands for human-computer interaction. The user interface has to consider the characteristics of mobility, for example, disturbing, random external forces because of the movement, parallel processing of orientation tasks, and so forth, for an optimal user input performance.

Subjective Workload: There are many definitions of the term *Workload*. One concise definition was given by O'Donnell & Eggemeier (1986): "… Workload refers to that portion of the operator's limited capacity actually required to perform a particular task." Subjective workload describes the effort invested by human operators into task performance. It can be assessed by subjective ratings.

Chapter L
Multilayered Approach to Evaluate Mobile User Interfaces

Maria de Fátima Queiroz Vieira Turnell
Universidade Federal de Campina Grande (UFCG), Brazil

José Eustáquio Rangel de Queiroz
Universidade Federal de Campina Grande (UFCG), Brazil

Danilo de Sousa Ferreira
Universidade Federal de Campina Grande (UFCG), Brazil

ABSTRACT

This chapter presents a method for the evaluation of user interfaces for mobile applications. The method is based upon an approach that combines user opinion, standard conformity assessment, and user performance measurement. It focuses on the evaluation settings and techniques employed in the evaluation process, while offering a comparison between the laboratory evaluation and field evaluation approaches. The method's presentation and the evaluation comparison will be supported by a discussion of the results obtained from the method's application to a case study involving a Personal Digital Assistant (PDA). This chapter argues that the experience gained from evaluating conventional user interfaces can be applied to the world of mobile technology.

INTRODUCTION

As proposals for new techniques and methods emerge for the evaluation of mobile device usability, it becomes more difficult for practitioners to choose among them. To be able to evaluate the efficacy of these techniques and methods, as well as to reproduce their steps, they have to be described in a level of detail not often found in the literature. Claims are often made without solid statistical results and are usually based on superficial descriptions. This makes it difficult, if not impossible, to compare alternative choices. Given the features of these new devices (such as mobility, restrictive resources for information input and output, and dynamic contexts of use), HCI specialists may question the efficacy of the methods, techniques, and settings already known to them from previous experiences. Thus, the major question that is addressed is whether it is possible to adapt the methods, techniques, and settings from previous evaluation experiences to this new class of devices, given their distinctive features.

The most frequent question raised in the vast majority of studies presented in the literature is whether to adopt a field approach or a laboratory approach. However, little is discussed in terms of which techniques are best suited for the specific evaluation target and its context of use. While this polemic subject may represent to the HCI specialist an import concern, it is equally important to consider the efficacy of the method, which accompanies this choice of approach (efficacy meaning the quality of the answers to the questions formulated as the basis of the evaluation). This is because the efforts employed in the evaluation may not pay off if a method is not well chosen or well employed.

This chapter presents a method for evaluating mobile devices based upon a set of techniques already known to the HCI specialist community. Each technique evaluates the problem from different perspectives: the user perspective (expressed as views on the product obtained through a questionnaire), the specialist's perspective (expressed when analyzing the user performance during the usability evaluation), and the usability community perspective (expressed in the form of standards conformity assessment). Each of these perspectives identifies evaluation problems and, when overlaid, they lead to a more reliable and complete product appraisal.

The remainder of this chapter is structured as follows. The second section gives a brief overview of the evaluation approaches currently in use for mobile devices, according to the literature review. The third section outlines the multi-layered approach. The fourth section illustrates the application of the multi-layered approach by means of a case study involving a Personal Digital Assistant (PDA). The fifth section discusses the results of the case study and their implications for the questions posed in this chapter. Finally, the sixth section concludes with the discussion of future trends in evaluation methods and how to apply the existing experience to the evaluation of this new class of products.

USER INTERFACE EVALUATION FOR MOBILE DEVICES

In the context of user-centered design processes, a significant portion of usability work involves the coordinated acquisition of valid and reliable data by a team of professionals. These specialists have varied backgrounds and skills and employ a number of evaluation methods. The expected result is an improved system design. This is achieved by the successful identification of a system's usability problems that might impact the interaction quality for a range of users.

Usability data consists of any information that can be used to measure or identify factors affecting the usability of a system being evaluated (Hilbert & Redmiles, 2000). These data are crucial for designing successful systems intended for human use. Such data are gathered by usability evaluation methods and techniques that can assign values to usability dimensions (Rosson & Carroll, 2002) and/or indicate usability deficiencies in a system (Hartson, Andre, & Williges, 2003). According to the International Organization for Standardization (ISO, 1998), usability dimensions are commonly taken to include user efficiency, effectiveness, and subjective satisfaction with a system in performing a specified task in a specified context.

Usability data are gathered via either analytic or empirical methods (Nielsen, 1993; Mayhew, 1999; Rosson & Carroll, 2002). Analytic methods, in which a system is evaluated based on its interface design attributes, are usually conducted by HCI specialists and do not involve human participants performing tasks. This means that these methods often rely on the specialists' judgment. Empirical methods, in which the system is evaluated based on observed performance in actual use, involve data collection of human usage.

Other classifications include direct methods (recording actual usage) and indirect methods (recording accounts of usage) (Holzinger, 2005). There are also formative and summative methods (Wixon & Wilson, 1997). The direct methods are used to generate new ideas and gather data during the development of a system in order to guide iterative design (Hix & Hartson, 1993). The indirect methods are used to evaluate existing systems and gather data to evaluate a completed system in use (Scriven, 1967). Discovery methods (also called qualitative methods) are used to discover how users work, behave, and think, and what problems they have. Decision methods (also called quantitative methods) are used in selecting a design among several alternatives or in picking elements of interface designs (Wixon & Wilson, 1997).

In essence, usability data have been classed in a number of other models and frameworks, often focusing on (1) the approach employed for gathering the data (including the resources expended and the degree of formality) (Danielson, 2006); (2) the context of use (including lighting, noise level, network connectivity, communication costs, communication bandwidth, and the social situation) (ISO, 1998; ISO, 1999; Jones & Marsden, 2006); (3) the nature and fidelity of the artifact being evaluated (EATMP, 2000); and (iv) the goal of the acquisition process (Kan, 2002).

It is a fact that usability evaluation for stationary computer systems has grown in the last two decades. In spite of debates still taking place within the HCI area, they are often based on a tacit understanding of basic concepts. One example of this understanding is in relation to the distinction between field and laboratory evaluation approaches and their importance to the area. Classical extensive guidelines were written that describe how usability evaluation in controlled environments should be conducted (e.g., Dumas & Reddish, 1999; Mayhew, 1999; Nielsen, 1993). Additionally, experimental evaluations of the relative strengths and weaknesses of different techniques are available that can be applied in a usability evaluation (e.g., Molich et al., 1998).

In the last decade, methodologies and approaches in HCI have been challenged by the increasing focus on systems for wearable, handheld, and mobile computing devices. One such move beyond office, home, and other stationary-use settings has pointed to the need for new approaches in designing and evaluating these systems (Kjeldskov, 2003). While the primarily task-centered evaluation approaches may be applicable to the desktop computing paradigm (often structured with relatively predictable tasks), they may not be directly applicable to the often-unpredictable continuous interaction possibilities and relatively unstable mobile settings. Additionally, it is not easy for evaluation methods to integrate completely or even adequately in real world or simulated settings contexts during the evaluation process. Authors argue that mobile computing demands not only real users but also a real or simulated context with device interaction tasks. It also demands real tasks or realistic task simulations.

There are a number of studies that discuss the question of whether the evaluation should be carried out in a laboratory or field context (e.g., Goodman et al., 2004; Kjeldskov & Stage, 2004; Kjeldskov et al., 2005; Po et al., 2004). All of these papers have a common theme, in that they apply a multi-method approach to performance measurement and discuss solutions for efficient data analysis. Nonetheless, it is important to note that the approach to usability evaluation depends on the relevance of the results presented as well as on the quality of the data analysis process. In general, the reports only present the results of the data analysis, omitting the details of the analysis process itself. While the data gathering method is critical for data quality, a more rigorous analysis on user comments and problem reports could help specialists better assess their choices.

There is a lot of current human-computer interaction research on alternatives for data collection methods and techniques. However, adequate data analysis and validation are only presented in few cases (e.g., Nielsen, 1994; Dumas & Redish, 1999; Po et al., 2004). In general, this aspect of the HCI research is poorly described in the literature, there being only vague conclusions and little guidance for attempts at successfully replicating the findings in other evaluation contexts. Many methods and techniques have been employed in the analysis of empirical data gathered during usability evaluations. Examples are for field testing analysis, video data analysis (Sanderson & Fisher, 1994), expert analysis (Molich et al., 1998), and head-mounted video and cued recall (Omodei et al., 2002). Its time-consuming character and its poor applicability for industrial purposes can explain the absence of an in-depth usage data analysis when under resource constraints (Baillie & Schatz, 2005). Nonetheless, it is strongly recommended for research purposes as a means to support new findings. For the same reason, it is equally important to provide sufficient detail to allow for replication and a substantiated choice of methods with similar levels of description.

THE MULTILAYERED EVALUATION APPROACH

The method described here was originally proposed for evaluating desktop interfaces. It was then adapted to evaluate the usability of mobile devices. It is based upon a multi-layered approach that combines standard conformity assessment, user performance measurement, and user satisfaction measurement. Each one of these evaluation techniques detects problems from a specific point of view. The multilayered approach is based on the premise that the combination of techniques (triangulation) will produce complementary and more robust results.

Standard Conformity Assessment

According to the International Organization for Standardization (ISO), conformity assessment means checking whether products, services, materials, processes, systems, and personnel measure up to the requirements of standards (ISO, 2006).

In its original version, this evaluation method adopts the standard ISO 9241 (*Ergonomic Requirements for Office Work with Visual Display Terminals*).

In the PDA case study it was found that only some parts of this standard can be applied to this mobile device: Parts 11 (ISO 9241-11, 1998), 14 (ISO 9241-14, 1997), 16 (ISO 9241-16, 1999), and 17 (ISO 9241-17, 1998). There are also some other standards that apply to this kind of device such as the ISO/IEC 14754 (*Pen-based Interfaces—Common gestures for text editing with pen-based systems*) (ISO/IEC 14754, 1999) and others that, although applicable to mobile devices, do not apply in this specific case. Examples are the ISO/IEC 18021 (*User interfaces for mobile tools for management of database communications in a client-server model*), since it is for devices capable of performing data interchange with servers (ISO/IEC 18021, 2002); and ITU-T E.161 (*Arrangement of digits, letters, and symbols on telephones and other devices that can be used for gaining access to a telephone network*, also known as ANSI T1.703-1995/1999, and ISO/IEC 9995-8:1994) (ITU, 2001).

User Satisfaction Measurement

User satisfaction has received considerable attention from researchers since the 1980s as an important surrogate measure of information systems success (Aladwani & Palvia, 2002; Goodhue & Thompson, 1995; Bailey & Pearson, 1983). While most user satisfaction measuring instruments were not Web-based at the time of development, others have been successfully validated in a Web-based environment (e.g., De Oliveira et al., 2005).

The user satisfaction diagnosis provides an insight into the level of user satisfaction with the product, highlighting the importance of the problems found and their impact on the product acceptance.

User Performance Measurement

The user performance measurement aims in general to provide data on the effectiveness and efficiency of a user's interaction with a product. It enables comparisons with similar products, or with previous versions of the same product along its development. Additionally, it can highlight areas where a product can be enhanced to improve usability. When used with the other methods, the evaluator can build a complete picture of the usability of a system.

The most significant user interface problems can be found by conducting experiments (usability tests) with representative users to observe how quickly, easily, and safely they can operate a product. The major change introduced in the original method concerns the introduction of field tests as a complement to the original laboratory tests.

The Experiment: Comparing Field and Laboratory Use of a PDA

The main objective of this study is to investigate the need for adapting the original evaluation method to the context of mobile devices, based on the analysis of the influence of the context (field versus laboratory and mobility versus stationary interaction) on the evaluation of mobile devices and applications.

The mobile device chosen as the target for this study was a PDA, the *Nokia 770 Internet Tablet* and some of its native applications. Tests were performed in a controlled environment (the usability laboratory) and also in the field. Twenty-four users took part in the experiment, divided into two groups of twelve.

Experiment Design

The study was designed to investigate the influence of the context (field and laboratory) and associated aspects such as mobility, settings, and so forth, and the user experience on the evaluation results. The independent variables are those that are not influenced by the context, by the test facilitator, or by external factors such as noise and lighting. An experiment plan was drawn from the study's objectives. The independent variables were chosen as follows:

Table 1. Plan for the experiment with the device Nokia 770

EXPERIMENT PLAN	
Target-Problems	1. With the shape/dimensions of the product 2. With the mechanisms for information input/output 3. With the processing power 4. With the navigation between functions 5. With information legibility
Test Objectives	1. Investigating the target problems 2. Detecting other problems
Objective Indicators	1. Task execution time 2. Number of incorrect actions 3. Number of incorrect choices 4. Number of repeated errors 5. Number of accesses to the online help 6. Number of off-line help (printed manuals) accesses
Subjective Indicators	1. Product ease of use 2. Task completion easiness 3. Input mechanism ease of use 4. Text input modes ease of use 5. Understandability of terms and labels 6. Understandability of messages 7. Help mechanism efficiency

- **Task context** comprises factors that may affect the users' behavior and their performance during the experiment (usability test). These factors may be internal or external to the user. The external factors originate in the field environment, examples being noise level and light intensity. The internal factors, on the other hand, are stress or other health conditions that may affect the user's mental and physical abilities.
- **User mobility** refers to the conditions under which the task is being performed. An example is if the user is required to work while being mobile, that is, moving between places or wandering while working.
- **User experience level** refers to the user's knowledge regarding mobile devices in particular and desktop computers systems in general.

The dependent variables are all dependant on the user's experience level:

- **Task time** represents the time taken by a device's user to perform a task.
- **Number of incorrect choices** measures how many times the user has made incorrect choices while selecting options in the interface through a menu dialogue.
- **Number of incorrect actions** measures how many times the same error (excluding the number of incorrect choices) was committed by the user while performing a task.
- **Number of accesses to the online help** and **number of accesses to the printed help** measure how many times the user accessed the online and printed help while performing a task.
- **Perceived usefulness** represents the user's opinion about the usefulness of the mobile application for the prescribed task.
- **Perceived ease of use** represents the user subjective satisfaction when using the mobile device.

Table 1 summarizes the experiment plan, which states the independent and dependent variables to be observed during the experiment and used as indicators to answer the research questions.

Test Environment

A software tool was used in the field environment to remotely capture the device's screen through a wireless connection to the lab. The user inputs (through keypad and stylus) were registered by a micro-camera coupled to the device and also remotely connected to the laboratory through a wireless connection. The interaction was registered in the controlled environment using two video cameras installed in the laboratory. One was focused on the users´ facial expressions and the other registered the device screen. As in the field environment, software was used to remotely capture the device's screen. Since the field setting required a wireless network, the field experiment was performed in the area surrounding the university's computer department. In both cases, the test facilitator was a human interface specialist who remained within reach in case the user required any explanation on the test procedure.

Table 2. User sample categorization

CATEGORY / CHARACTERISTIC	Beginner	Intermediate	Expert
Previous Computer Knowledge	Basic/ Intermediate	Intermediate/ Advanced	Intermediate/ Advanced
Previous Experience with *Nokia 770*	No	No	Yes

Participants

Users participating in the PDA experiment were selected on the basis of having previous experience with mobile devices (such as mobile phones), computers, and the Internet. They were also required to have some familiarity with the English language, since this is the language adopted in the device's user interface and in its documentation. The user sample was then classed according to the users' experience level into the categories shown in Table 2.

The recruited users were divided into two groups of 12 to participate in the field and laboratory tests. Based on user experience level, both groups were then subdivided into three subgroups of four beginners, four intermediates and four experts.

Materials

Laboratory Test Materials

- **Hardware:** The Nokia 770 Internet Tablet; PC based Workstation (2); Video cameras (3); Microphones (2).
- **Software:** VNC (Virtual Network Computing) software to capture the screens during the interaction with the device; the WebQuest tool with the questionnaires pre-test (to gather the user profile) and post-test (to collect and process the user satisfaction level).
- **Miscellaneous:** The Nokia 770 Internet Tablet Manual; chronometer (1); CDs for video backup; participant registration form; test conditions acceptance forms on which the users declared their acceptance of the experiment conditions; task script that consists of a written task description to guide the user during the session (versions for the user and for the evaluator); Form for event log.

Field Test Materials

- **Hardware:** The Nokia 770 Internet Tablet; PC-based Portable (laptop) Workstation (1); wireless video micro-camera (1); apparatus to support the video micro-camera (1); television set (1); VCR equipment (1).
- **Software:** VNC (Virtual Network Computing) software to capture the screens during the interaction with the device; WebQuest tool with the questionnaires pre-test (to gather the user profile) and post-test (to collect and process the user satisfaction level).
- **Miscellaneous***:* Chronometer (1); CDs and VHS tapes for video backup; participant registration form; test conditions acceptance forms on which the users declared to accept the experiment conditions; task script that consists of a written task description to guide the user during the session (versions for the user and for the evaluator); form for event log.

Camera Apparatus

The apparatus shown in Figure 1 was built to couple a video micro-camera to the mobile device. This allowed the recording of user interaction through a remote link with the laboratory computer.

Figure 1. Apparatus to support video camera during experiment

The WebQuest Tool

A Web tool named *WebQuest* supports the method application (De Oliveira et al., 2005). This tool was developed to support the specialist during data collection, to provide automatic score computation, to perform statistical analysis, and to generate graphical results. *WebQuest* also enables the specialist to reach a more diverse and geographically widespread sample of users through the Internet. One of its features is a flexible questionnaire structure, which enables specific context adaptation and, by means of an estimation model, ensures a higher degree of confidence on the indicators of user satisfaction. Currently *WebQuest* supports two questionnaires: (i) a pre-test questionnaire, the *USer* (*User Sketcher*), conceived to raise the profile of the system users; and (ii) a post-test questionnaire, the *USE* (*User Satisfaction Enquirer*), conceived to raise the user degree of satisfaction with the system. The pre-test questionnaire incorporates a model to estimate the user's subjective satisfaction and can be answered directly on the Web. The questions are related to the users´ physical characteristics, knowledge, and skills. Both questions and answers are configurable.

As for the *USE* (*User Satisfaction Enquirer*), it allows gathering quantifiable variables on the user acceptance of the device. Three of its aspects are of special interest. Firstly, it incorporates a model to estimate user subjective satisfaction. Secondly, the questionnaires can be answered directly on the Web. Thirdly, the items are partially or totally configurable. The adoption of an estimation model by *USE* allowed us to establish a subjective satisfaction coefficient directly from the inspection of the respondents' samples. The *WebQuest* tool allows the specialist to easily edit the questionnaire's items. These items are organized into groups: (1) *fixed*, which are applicable to various evaluation contexts and thus are not allowed to be altered; (2) *semi configurable*, which allow for changes in the answer options; and (3) *configurable*, which can be fully configured (both the question and respective options of answers). *USE* supports the specialist from the data collection through to automatic score computation, performing statistical analysis, and generating graphics with the results.

Experiment Procedure

The techniques employed in the experiment procedure were the observation and subsequent video analysis for accumulating quantitative data (such as time spent and error rate). An automated video capturing tool recorded the interactions of the subjects during the field tests to ensure a non-intrusive observation method. During task execution, the users were asked for their consent before being filmed. The conditions of test-subject participation included a written commitment not to disclose any product information. The users were also asked to give consent so that their images and/or sound recordings made during the experiment could be used for research purposes or in a multimedia product evaluation report. On the other hand, the users were given assurances from the evaluation team that no personal information or individual performance results would be disclosed.

The first step in following the method consisted in defining the evaluation scope for the product as well as a scenario for the test. Table 3 illustrates the sequence of tasks performed during the experiment.

The decision was based on a heuristic evaluation performed by the evaluation team. This initial step also supports the definition of a general

Table 3. Test scenario and sequence of tasks to be performed during experiment

TASKS IN SCRIPT	
T01	**Initializing the device**
T02	**Searching for books in an online store**
T03	**Visualizing a PDF file**
T04	**Entering textual information**
T05	**Using the electronic mail**
T06	**Using the audio player**

profile for the user sample and a classification into categories. Following, the method the evaluation objectives were defined. These became the basis for choosing the product evaluation scenario (product context of use and laboratory settings) and the corresponding tasks to be performed by the users during the experiment. Having planned the evaluation, a pilot test was conducted to verify the adequacy of the proposed experiment procedural, materials, and environment. Through this fine tuning procedure it was found, in the PDA case study, that the time to perform the tasks had been underestimated. This resulted in re-dimensioning the test scenario to six tasks, with a review of the tasks themselves to fit the established session time of sixty minutes to prevent user tiredness.

All subjects were submitted to the same procedure prescribed in the experiment protocol. The study was conducted first in a laboratory setting and then in the field environment. During the field tests the participants were taken outdoors, and the tasks were conducted in an environment that was as close to real-use conditions as possible.

The experiment conducted in the usability laboratory had the audio and video of each session recorded. In the field experiment, only the video of the sessions was recorded, supplemented by comments written by the specialist. As described in the experiment protocol, each session consisted of the following steps: (1) introducing the user to the test environment by explaining the test purpose, the procedure to be followed and the ethics involved in terms of the conditions of participation; (2) applying the pre-test questionnaire; (3) performing the task script; (4) applying the post-test questionnaire; and (5) performing a non-structured interview.

At the time of the experiment, the *Nokia 770 Internet Tablet* device was not yet widely known in the Brazilian market. The users who claimed to have had no previous contact with it were given a quick introduction. This introduction consisted of an instructional material given to the recruited users and also a quick explanation about the device's input and output modes and its main resources.

Results

The results obtained from the experiment in which the multi-layered method was applied support the original assumption that, in spite of the distinctive features of this class of devices, it is possible to adapt from the evaluation experience with conventional devices. This conclusion is supported by the evidence that the evaluation context did not significantly influence the user performance or the opinion about the device's usability, given through the analysis of the objective and subjective indicators associated with the experiment.

Standard Conformity Assessment Results

The results of the conformity assessment to the standards ISO 9241 Parts 14 and 16 and ISO 14754 are illustrated in Table 4. According to ISO, con-

Table 4. Nokia 770 conformity assessment with standards

Standard	#Sar	#Ar	AR (%)
ISO 9241 Part 14	45,0	53,0	84,9
ISO 9241 Part 16	26,0	33,0	78,8
ISO 14754	4,0	11,0	36,4

Sar—Successfully adhered recommendations
Ar—Applicable recommendations
AR—Adherence Rate

$$AR = \frac{Sar}{Ar} x100\%$$

formity assessment results can be summarized by computing an *adherence rate* (AR). This is the percentage of the applicable recommendations (Ar) that were successfully adhered to (Sar).

In spite of the device's characteristics that limit the number of applicable recommendations, these results corroborate the idea that the standards inspection is still applicable in the evaluation process. The efficacy of this technique can be considerably improved if it is based upon standards conceived specifically for mobile devices, which could evidence more usability problems.

User Satisfaction Measurement Results

For the PDA case study context, both questions and answers of the *USE* questionnaire were configured. The questionnaire was applied soon after the usability test and answered using the mobile device itself. As mentioned before, its purpose was to collect information on the user's degree of satisfaction with the device and on aspects such as interface navigation, documentation, and overall impressions.

The *USE* was composed of three sections. The first section is relative to "the product Use and Navigation." It is composed of 17 items and focuses on aspects such as menu items, navigation between functions, understandability of the messages, ease of use of the basic functionalities, and of the device's input and output mechanisms. The second section consists of six questions related to the online and off-line (printed manuals) documentation. The last section ("You and the product") consists of 15 items and aims to get the user's impressions and product acceptance level. The first 23 items use a 5-point semantic scale (1: *very easy*; 2: *easy*; 3: *not easy nor difficult*; 4: *difficult*; and 5: *very difficult*). The last 15 items use another 5-point semantic scale (1: completely agree; 2: agree; 3: do not agree nor disagree; 4: disagree; and 5: completely disagree). The users were asked to answer the questions and to assign an importance level to each one of them, on a scale from 0 to 10.

For the post-test questionnaire, *USE* adopts the model proposed by Bailey and Pearson (Bailey & Pearson, 1983) for measuring the overall user's sense of satisfaction. The following adaptations to the dimensions were considered: (1) the association of only one (1) semantic differential scale to the items, instead of the four (4) semantic differential scales, as proposed in the original model; (2) the adoption of a 5-point Likert scale, delimited by the ends -2 and 2 (instead of the 7-point scales delimited by the ends -3 and 3 as originally proposed); and (3) the incorporation of a 11-point importance scale (0 corresponding to *non applicable*), varying from 0.0 to 1.0 in intervals of 0.1 (instead of the original 7-point scales, which varied from 0.1 to 1.0 in intervals of 0.15).

The user's subjective satisfaction indicators for the PDA case study were 0.330 for the laboratory experiment and 0.237 for the field experiment. The normalized value ranges of the user satisfaction concerning a product are 0.67 to 1.00 (*Extremely Satisfied*), 0.33 to 0.66 (*Very satisfied*), 0.01 to 0.32 (*Fairly satisfied*), 0.00 (*Neither satisfied nor unsatisfied*), 0.01 to 0.32 (*Fairly dissatisfied*), 0.33 to 0.66 (*Very dissatisfied*), and 0.67 to 1.00 (*Extremely dissatisfied*). This is in accordance with the Bailey and Pearson model (Bailey & Pearson, 1983). The results obtained correspond respectively to *Very satisfied* and *Fairly satisfied*.

Performance Measurement Results

The User Sample Profile

The user sample profile was drawn with the support of the questionnaire *USer*. It was composed of 13 male and 11 female users, of which eight were *undergraduate students*, 12 *post-graduate students*, two *graduate level*, and two *post-graduate level*. The ages varied between *18* and *29* years. They were mainly *right handed* and mostly used some sort of reading aid (either *glasses* or *contact lenses).* All of them had at least one year of *previous experience of computer systems* and were currently using computers on a *daily* basis.

User Performance Data Analysis

After having analyzed the data gathered during the experiment on the user performance and having analyzed the list of problems found with this technique, it was possible to evaluate their impact and class them as: minor (50%), medium (50%), major (0%), consistency (35.7%), recurrent (64.3%), and general (0%).

The data analysis consisted of a statistical processing and had two main purposes: (1) to investigate the influence of the context on the results of the evaluation method (through the comparison of the results obtained from both environments); and (2) to investigate the influence of the user experience with the mobile device on the test results within each context. For the latter purpose, the three categories illustrated in Table 2 were used.

The statistic analysis performed consisted of: (1) building a report with univariance statistics; (2) generating the covariance matrices for the objective and subjective indicators that were previously defined; (3) applying the one-way F ANOVA test (Tabachnick & Fidell, 2006) to the data obtained from the previous step in order to investigate possible differences; and (4) applying the Tukey-Kramer process (Tabachnick & Fidell, 2006) to the one-way F ANOVA results aiming to investigate if the found differences were statistically significant to support inferences from the selected sample. The result of this technique was the identification of 13 problems, of which 92.3% were found in the laboratory and 61.5% in the field as: Laboratory (38.5%); Field (7.7%); and Laboratory & Field (53.8%).

Overlaying Results

Since the multi-layered evaluation is based upon a triangulation of results, Table 5 summarizes the usability problem categories identified by the three techniques.

The numbers correspond to the identification of each problem from a list of problems found through each technique. As can be seen from Table 5, some of the usability problem categories were more related to the performance measurement (e.g., hardware aspects, help mechanisms, processing capacity) whereas others (e.g., menu navigation, presentation of menu options) were identified by the conformity assessment. It was possible to identify 66.7% of the problems found by other methods when combining the results from the post-test questionnaire with the user comments made during the experiment and the informal interview at the end of the experiment. This confirms the importance of combining techniques to obtain a more complete result when performing usability evaluation. It must be pointed out that 29.62% of the problems based on the user opinion about the product were in disagreement with the results of the other two evaluation dimensions (specialist and the community points of view). This discrepancy can originate from the users' perception of product quality and the perception of their own skills to perform the task, accepting full responsibility over the difficulties that might arise during the interaction. When overlaying the problems in Table 5, in the category *Menu navigation*, the same problem was found by the techniques Standards Inspection and Performance Measurement.

DISCUSSION

From this study's data analysis it became evident that certain problem categories are better found by specific techniques, as shown in Table 5. For instance, problems associated to the device's physical characteristics are better found by means of conformity assessment, whereas the user performance located problems associated to the device's applications.

The analysis of the pre-test and post-test questionnaires and the informal interviews showed that domain knowledge and computer literacy have significant influence on user performance with mobile devices. This was true both under laboratory conditions and in the field, in relation to the incidence of errors. The univariate analyses of variance of the performance variables: *Time*, *Errors*, and *Accesses to help, are* presented in Table 6.

Table 5. Overlay of results obtained with the three evaluation techniques

PROBLEM CATEGORY	SI	PM	SM
Location and sequence of menu options	✓ (05)		× (05)
Menu navigation	✓ (02)	✓ (01)	
Presentation of menu options	✓ (02)		
Information feedback	✓ (01)		
Object manipulation	✓ (05)		
Symbols and icons	✓ (02)		× (02)
Text entry via stylus (Writing recognition)	✓(07)	✓ (01)	✓ (08)
Text entry via virtual keyboard		✓ (01)	✓ (01)
Processing power		✓ (02)	✓ (02)
Hardware issues		✓ (03)	✓ (03)
Fluent tasks execution		✓ (05)	✓ (05)
Online and offline help		✓(01)	✓ (01)

Legend:

SI—Standards Inspection
PM—Performance Measurement
SM—Satisfaction Measurement

× - Contradictory findings
✓ - Consistent findings

Table 6. Influence of the user experience on the performance indicators: Time, Number of errors, and accesses to help

Independent variable	Dependent variable	p-Value (Lab)	p-Value (Field)	Significance (α=0.05)
Experience	Task Time	0.081	0.081	Not significant
Experience	Errors	0.011	0.002	Significant
Experience	Help Accesses	0.427	-	Not significant

From this table, it can be seen that the user experience level had a more significant effect on the number of errors in the field experiment than in laboratory experiment.

The studies in the literature fit basically into two categories: (1) user mobility, which means moving while using the device (inside of a laboratory or outdoors) and (2) user attention division. However, this study considers both aspects as part of the task context. In this experiment, the field test subjects were free to choose between moving or remaining still as they performed the task with the mobile device. During the informal interview the users stated that in a real context they would not perform the experiment tasks on the move**,** since they demanded too much attention. The specialist encouraged users to wander around the environment, although they could choose to enter a room in the building, sit down, or even lay the device on a table (which they did in most cases, under the argument that this setting was more comfortable). The movement registered was limited to situations in which the user waited for some device processing. (e.g., Web page downloads). There was a clear interference of the environment on the user attention during the field tests while moving.

The device's physical characteristics affected the user performance and the data gathering during the experiment**.** Outdoors, in ambient light, the device's legibility was reduced and aggravated by the reflections on the screen. According to the user's opinion stated during the informal interview, the camera apparatus did not interfere with the task execution, but the majority decided to lay the device down during task execution.

As for the entry of text information, the users showed a preference for the virtual keyboard instead of hand written character recognition. Based on their comments, as well as on the informal interview, it was concluded that writing long messages is very cumbersome both using the virtual keyboard and using the handwriting recognition application. Confirming previous findings, the experiment demonstrated that applications that require a lot of interaction and user attention are inappropriate for performing while walking due to attention division. This conclusion reinforces that, for the device targeted in this study, in spite of its mobility, the evaluation settings did not need to differ substantially from the one employed in the evaluation of stationary devices since the users tend not to wander while performing tasks that demand their attention or consisted of text input.

Until recently, studies have been published which deal with new paradigms and evaluation techniques for mobile devices. Few of the proposed new techniques are really innovative if compared to the ones traditionally employed. On the other hand, the main argument for proposing new techniques concerns the user and device mobility and the influence of this mobility on user performance. In contrast, this study evaluated the effect of mobility not only from the user performance perspective but also from user opinion point of view and the user level of satisfaction. From the application of the multi-layered approach, the data gathered and analyzed support the initial assumption that minor adaptations in the traditional evaluation techniques and respective settings are adequate to accommodate the evaluation of the category of mobile devices targeted by this study.

The conclusions corroborate with the views of the authors and that of Po (Po, 2003) that the laboratory and field evaluations do not diverge but are complimentary. As shown in this study, they both add to the evaluation process, producing data that is significant to the process and reinforcing the relevance of a multi-layered approach for the usability evaluation of mobile devices.

FUTURE TRENDS

Mobile devices impose challenges to the usability evaluation that are unique in respect to the observation strategies and the conception of test scenarios. With the continuous technological advances, a wider variety of new devices is being released into the market, challenging users with the complexity of the interaction. In this scenario, the importance of the product usability is undisputable as is also the correct choice of evaluation methods, techniques, and tools.

One emerging trend in the mobile devices evaluation field is the possibility of gathering data in an unobtrusive way, using tools for remote, and automatic data capture that are transparent to the user. Developing those tools is a challenging activity given the inherent restrictions presented by the mobile devices (such as their limited processing power and limited storage capacity). But, in spite of the current limitations, it was shown in this study that the tools are becoming available to provide a great contribution to the evaluation setup and that these tools would benefit from further development.

REFERENCES

Aladwani, A., & Palvia, P. (2002). Developing and validating an instrument for measuring user-perceived Web quality. *Information & Management*, *39*, 467-476.

Bailey, James E., & Pearson, S. W. (1983). Development of a tool for measuring and analyzing computer user satisfaction. *Management Science*, *29*(5), 530-545.

Baillie, L., & Schatz, R. (2005). Exploring multimodality in the laboratory and the field. In *Proceedings of the 7th International Conference on Multimodal Interfaces* (pp. 100–107).

Danielson, D. R. (2006). Usability data quality. In C. Ghaoui (Ed.), *Encyclopedia of human-computer interaction* (pp. 661-667). Hershey, PA: Idea Group Reference.

De Oliveira, R. C. L., De Queiroz, J. E. R., Vieira Turnell, M. F. Q. (2005). WebQuest: A configurable Web tool to prospect the user profile and user subjective satisfaction. In G. Salvendy (Ed.), *Proceedings of the 2005 Human-Computer Interaction Conference. (The management of information: E-business, the Web, and mobile computing)* (Vol. 2) Nevada: Lawrence Erlbaum Associates (U.S. CD-ROM Multi Platform).

Dumas, J. S., & Redish, J. C. (1999). A practical guide to usability testing (revised ed.). Exeter, UK: Intellect.

EATMP. (2000). *Human factors integration in future ATM systems—Methods and tools* (Tech. Rep. HRS/HSP-003-REP-03). European Organization for the Safety of Air Navigation—European Air Traffic Management Programme. Retrieved August 13, 2006, from http://www.eurocontrol.int/humanfactors/gallery/content/public/docs.

Goodhue, D. L., & Thompson, R. L. (1995). Task-technology fit and individual performance. *MIS Quarterly*, *19*(2), 213-236.

Goodman, J., Brewster, S., & Gray, P. (2004). Using field experiments to evaluate mobile guides. In *Proceedings of 3rd Annual Workshop on HCI in Mobile Guides* (pp. 1533-1536).

Hartson, H. R., Andre, T. S., & Williges, R. C. (2003). Criteria for evaluating usability evaluation methods. *International Journal of Human-Computer Interaction, 15*(1), 145-181.

Hilbert, D. M., & Redmiles, D. F. (2000). Extracting usability information from user interface events. *ACM Computing Surveys*, *32*(4), 384-421.

Hix, D., & Hartson, H. R. (1993). *Developing user interfaces: Ensuring usability through product & process*. New York: John Wiley and Sons, Inc.

Holzinger, A. (2005). Usability engineering methods (UEMs) for software developers. *Communications of the ACM*, *48*(1), 71-74.

ISO 9241-11. (1998). *Ergonomic requirements for office work with visual display terminals (VDTs)—Part 11: Guidance on usability*. International Organization for Standardization, Geneva, Switzerland.

ISO 9241-14. (1997). *Ergonomic requirements for office work with visual display terminals (VDTs)—Part 14: Menu dialogues*. International Organization for Standardization, Geneva, Switzerland.

IS09241-16. (1999). *Ergonomic requirements for office work with visual display terminals (VDTs)—Part 16: Direct manipulation dialogues*. International Organization for Standardization, Geneva, Switzerland.

IS09241-17. (1998). *Ergonomic requirements for office work with visual display terminals (VDTs)—Part 17: Form filling dialogues*. International Organization for Standardization, Geneva, Switzerland.

ISO 13407. (1999). *Human-centered design processes for interactive systems.* International Organization for Standardization, Geneva, Switzerland.

ISO. (2006). *ISO and conformity assessment*. Retrieved September 23, 2006, from http://www.iso.org/iso/en/prods-services/otherpubs/pdf/casco_2005-en.pdf.

ISO/IEC 14754. (1999). Information technology—pen-based interfaces—common gestures for text editing with pen-based systems. International Organization for Standardization, Geneva, Switzerland.

ISO/IEC 18021. (2002). *Information technology—user interfaces for mobile tools for management of database communications in a client-server model*. International Organization for Standardization, Geneva, Switzerland.

ITU-T E.161. (2001). *Arrangement of digits, letters and symbols on telephones and other devices that can be used for gaining access to a telephone network*. International Telecommunications Union-telecommunications, Geneva, Switzerland.

Jones, M., Marsden, G. (2006). *Mobile interaction design*. Chichester, West Sussex: John Wiley and Sons, Inc.

Kan, S. H. (2002). *Metrics and models in software quality engineerin*g (2nd ed.). Reading, MA: Addison-Wesley Professional.

Kjeldskov, J., Graham, C., Pedell, S., Vetere, F., Howard, S., Balbo, S., & Davies, J. (2005). Evaluating the usability of a mobile guide: The influence of location, participants and resources. *Behavior and Information Technology*, *24*(1), 51–65.

Kjeldskov, J., & Stage, J. (2004). New techniques for usability evaluation of mobile systems. *International Journal on Human and Computer Studies*, *60*(5-6), 599–620.

Mayhew, D. J. (1999). *The usability engineering lifecycle*. San Francisco: Morgan Kaufmann Publishers Inc.

Molich, R., Bevan, N., Curson, I., Butler, S., Kindlund, E., Miller, D., et al. (1998). Comparative evaluation of usability tests. In *Proceedings of the Usability Professionals Association Conference* (pp. 189-200).

Nielsen, J. (1993). *Usability engineering.* Boston: Academic Press.

Omodei, M. A., Wearing, J., & McLennan, J. P. (2002). Head-mounted video and cued recall: A minimally reactive methodology for understanding, detecting and preventing error in the control of complex systems. In *Proceedings of the 21st European Annual Conference of Human Decision Making and Control.*

Po, S., Howard, S., Vetere, F., & Skov, M. B. (2004). Heuristic evaluation and mobile usability: Bridging the realism gap. In *Proceedings of the Mobile HCI* (pp. 49–60).

Rosson, M. B., & Carroll, J. M. (2002). *Usability engineering: Scenario-based development of human-computer interaction*. San Diego, CA: Academic Press.

Sanderson, P., & Fisher, C. (1994). Usability testing of mobile applications: A comparison between laboratory and field testing. *Human-Computer Interaction*, *9*, 251–317.

Scriven, M. (1967). The methodology of evaluation. In R. W. Tyler, R. M. Gagne, & M. Scriven (Eds.), *Perspectives in curriculum evaluation* (pp. 39- 83). Skokie, IL: Rand McNally.

Tabachnick, B. G., & Fidell, L. S. (2006). *Experimental designs using ANOVA* (1st ed.). Duxbury Applied Series. Pacific Grove, CA: Duxbury Press.

Wixon, D., & Wilson, C. (1997). The usability engineering framework for product design and evaluation. In M. Helander, T. K. Landauer, & P. Prabhu (Eds.), *Handbook of human-computer interaction* (2nd ed.) (pp. 653-688). New York: John Wiley and Sons, Inc.

KEY TERMS

Conformity Assessment: A collective term used for a number of techniques used to determine if a product, system, or process (including design) meets a defined specification.

Device Mobility during a Usability Evaluation: The ability to interact with the user and continue to perform its functions while being transported.

Efficacy of an Evaluation Method or Technique: Translated into the number of problems found, gravity of those problems versus the time, and cost of performing the experiments.

Likert Scale: An attitude scale in which respondents indicate their degree of agreement/disagreement with a given proposition concerning some object, aspect, person, or situation.

Multi-Layered Evaluation Approach: A product or prototype usability evaluation method that combines techniques for data gathering and analysis based on multiple perspectives (the user's, the specialist's, and the usability community). The results are overlaid in order to find discrepancies and offer more robust results.

User Mobility during the Usability Evaluation: The ability to move while performing a task with a product.

User Performance Measurement: The process of gathering actual data from users as they work with a system and its documentation. Usually, the user is given a set of tasks to complete and the evaluator measures the relevant parameters such as the percentage of tasks or subtasks successfully completed, time required to perform each task or subtask, frequency and type of errors, duration of pauses, indications of user frustration, and the ways in which the user seeks assistance.

User Satisfaction Measurement: The process of obtaining qualitative and quantitative information which indicates the extent to which user expectations concerning some object, process, product, or situation are being met. Such information can be obtained in a variety of ways, both formally and informally.

Virtual Network Computing (*VNC*): A desktop sharing system that uses the RFB (Remote Frame Buffer) protocol to remotely control another computer. It transmits the keyboard presses and mouse clicks from one computer to another over a network, relaying the screen updates back in the other direction.

Chapter LI
Theory and Application of the Privacy Regulation Model

Jaakko T. Lehikoinen
Nokia Research Center, Finland

ABSTRACT

Privacy is one of the most essential topics to be investigated when assessing user acceptance of new applications and services enabling disclosure of personal information. Mobility increases the demand on taking privacy into consideration when designing and developing these kinds of systems. This chapter presents a privacy management model, which facilitates evaluation of privacy aspects of communication technology. The applicability of the model is tested in a field trial that was carried out to assess user acceptance of a mobile social awareness system. Gathered evidence shows that the model helps researchers and designers to deal with privacy aspects of mobile technologies.

INTRODUCTION

Privacy is often seen as a design problem in the fields of human-computer interaction (HCI), computer supported cooperative work (CSCW), and ubiquitous computing (ubicomp). Privacy and its regulation play a significant role in defining acceptability of new communication technological solutions. However, user-perceived privacy threats may not correspond with the real risks related to the technology. For instance, Adams' (1999 and 2000) studies point out that new technologies are often considered acceptable if the invasion of privacy is not personally faced, even if the technology has major *potential* privacy risks.

In a mobile domain, privacy as a design factor is even more important than in stationary contexts of use. Mobile devices, such as mobile phones, are carried along throughout the day in a variety of situations. This has itself caused both threats and possibilities in personal privacy regulation. For example, receiving a phone call in a crowded situation may be inconvenient for the recipient due to the fact that one needs to carefully think about how to filter words in the ongoing conversation. On the other hand, mobility makes it possible to go to a peaceful place to talk with the other party. In addition, text messaging allows a user to have extremely private dialog regardless of a context. However, in many ubicomp scenarios, privacy is a real concern.

Before continuing privacy needs to be defined. It is a complex concept and can be understood only in relation to the surrounding social world. Philosophers and social scientists have produced different kinds of definitions of privacy. Schoeman (1992) has distinguished three most typical definitions. Firstly, privacy can be identified as the measure of control an individual has over information about himself, intimacies of personal identity, and/or who has sensory access to him. Secondly, privacy can be understood as a state or condition of limited access to a person. Thirdly, privacy can be thought as a claim of individuals to decide about dissemination of information concerning themselves. One of the most frequently referred philosophers, Alan Westin (1967), extends the latter definition by saying "Privacy is the claim of individuals to determine for themselves when, how and to what extend information about themselves is communicated to others."

Privacy can also be defined as a border between society and one's personal affairs (Soppera & Burbridge, 2004). Marx (2001) uses the following border categories; natural borders (governed by senses and physical boundaries), social borders (the expectation that information is shared within a social group), spatial or temporal borders (separate aspects of one's life), and ephemeral borders (based on the assumption that information is not preserved longer than expected). He also argues that the terms public and private are more ambiguous than what is generally understood.

Altman (1975), whose theory is used as a foundation for the privacy regulation model presented in this chapter, brings yet another definition of privacy; he understands privacy as a dialectic and dynamic boundary regulation process. In interpersonal relations, especially in face-to-face settings, the dynamic boundary regulation process means that an individual manages social interaction and personal privacy through different behavioral mechanisms, such as verbal behavior (content of speech), paraverbal behavior (voice intensity), personal spacing (distance from others), and territorial responses (personalizing and controlling geographical areas and objects). Depending on the circumstance, an individual uses these mechanisms in different ways; one mechanism may substitute the other according to changes in the social context. Altman also states that behavior may change in a process of time, thus the process of privacy regulation is dynamic.

Altman (1975) uses the concepts of *desired privacy*, *interpersonal control mechanisms*, and *achieved privacy* to illustrate privacy regulation process. Achieved privacy represents the actual amount of interaction with others and it may be more or less than the desired privacy, or it can match it. An optimal level of privacy exists, when the achieved privacy equals the desired privacy. When the actual amount of interaction is less than desired, Altman talks about *isolation* and when it is more, he uses the term *crowding*. The terms *inputs* and *outputs* describe people's behavior in a social situation. For instance, listening to other's conversation represents inputs from the others and attending actively to a discussion, and presenting one's own views to the subject matter represent the outputs from self to others. To conclude, in a state of desired privacy the inputs and outputs are in a level that an individual wishes.

Privacy is often seen as a design problem in the arena of communication technological research and development. Palen and Dourish (2003) bring out two essential factors affecting privacy management; *recordability* and *unawareness of the final audience*. The first factor refers to the fact that mediated communication can be recorded, which means that the mediated information can be retrieved and re-used after the information disclosure. This makes the control of personal information difficult or even impossible once it has been disclosed. Recordability and the fact that mediated communication is, in most cases, taking place in such environments where the communication parties cannot see each other, cause unawareness of the final audience. One of the most typical examples addressing these two cases is forwarding someone's e-mail without asking permission from the original sender.

These two factors are relevant in basic communication applications and services but when designing or evaluating more sophisticated communication solutions, other factors need to be taken into consideration as well. This chapter builds a holistic approach to privacy regulation, and thus provides a theoretical tool for designing and evaluating novel mobile communication systems. The privacy regulation model aims at presenting the elements affecting users' perceptions of information disclosure in computer mediated communication.

The next section presents earlier research on privacy regulation in the fields of ubicomp and HCI. After that, the privacy regulation model is presented and it is followed by an example case of applying the model to evaluating a novel mobile communication solution. The chapter concludes with a summary of the key aspects of privacy regulation and presents directions for further research.

PREVIOUS PRIVACY RESEARCH

Privacy has been studied fairly extensively in the fields of HCI and ubicomp. Many studies state that a lack of privacy is a real concern in ubiquitous computing environments (e.g., Al-Muhtadi, Campbell, Kapadia, Mickunas, & Seung , 2002; Soppera & Burbridge, 2004). For instance, Langheinrich (2002) argued that perfect privacy protection cannot be achieved in ubicomp. However, he says that the systems should be designed in such a way that they help others respect our personal privacy, enable us to be aware of our privacy, and rely on social and legal norms to protect us from wrongdoers.

Several models and frameworks related to privacy, like risk models or privacy control models, are presented in the earlier HCI and ubicomp research. Many of them are built to tackle a certain context of use or a field of technology, such as multimedia communication, location services, or proximity technologies. For instance, Bellotti and Sellen (1993) investigate the privacy threats in the context of ubiquitous computing in the working environment and continue by presenting a framework for designing privacy in ubicomp environments. According to Bellotti and Sellen, privacy threats can be classified into two categories: technological and user interface design related threats that are coupled with social behavior. They bring out two essential aspects of the design for privacy: *control* and *feedback*. The design framework proposes that a system should provide control and feedback for at least the following system and user behaviors: capture (information that is acquired), construction (how the information is processed), accessibility (who gets access to the information), and purposes (what is the information used for).

The works of Adams and Sasse (1999) and Adams (2000) are along similar lines to Bellotti and Sellen (1993). They analyze privacy in networked and ubiquitous multimedia communications, and mechanisms and policies for protecting users' privacy in ubiquitous multimedia applications. They have identified four aspects of privacy: information receiver, information sensitivity, information usage, and the context of disclosure. Similar aspects can also be found from the work of Soppera and Burbridge (2004), which addresses the following issues: data and identity, data collection and data usage, storage, and access. In contrast to earlier research, this work concludes that anonymity is a key enabler in maintaining privacy in ubicomp.

The risk model proposed by Hong, Ng, and Lederer (2004) provides a tool for identifying concrete privacy issues in ubicomp and prioritizing them. The model consists of analysis and management parts. The analysis part consists of social and organizational issues (such as the users and the different kinds of information involved), as well as the technology. Further, it considers the relationship between data sharers and observers. The management part helps in organizing and prioritizing the issues gathered in the analysis and identifying solutions.

Lederer, Dey, and Mankoff (2002) presented a conceptual model of what they call "everyday privacy," described as end-users' exposure to, and control over, personal information collection. The model is based on synthesis of Lessig's societal-scale model (2005) and Adam's perceptual model (1999). In the synthesized model, legal, market, normative and architectural forces, combined with contextual factors, constrain the possible levels of privacy. Within the constrained range, a user's subjective values determine the actual level of preferred privacy.

Regarding access to personal data presented in many of the privacy models, Jiang, Hong, and Landay (2002) have proposed a principle of minimum asymmetry. Asymmetric information is related to situations in which some actor has private information that is relevant for everyone. By applying the principle of minimum asymmetry, privacy-aware ubicomp physical spaces and systems should decrease the asymmetry of information between data owners, data collectors, and data users.

Price, Adam, and Nuseibeh (2005) have approached privacy from a different angle in their model for user control over privacy. They have identified four layers related to privacy in ubicomp: the regulatory regime a user is currently within, the type of ubicomp service required, the type of data being disclosed, and the personal privacy policy being applied. The model balances the user's privacy preferences with privacy regulations and provides means to protect location privacy. This work also discusses balancing trade-offs between privacy and receiving ubicomp services.

Field studies on ubicomp privacy have also been conducted. For instance, Beckwith (2003) has studied privacy in an elderly care facility in an established environment. In his study, he used a model loosely based on Adam's and Sasse's findings (1999). Many issues came up, such as the fact that people sometimes forgot they were monitored, which affected their control over information disclosure.

To conclude, as Soppera and Burbridge (2004) state, "the examination of privacy in the area of pervasive computing is immature." This argument applies generally to the field of designing and evaluating user interfaces for mobile communication technologies.

THE PRIVACY REGULATION MODEL

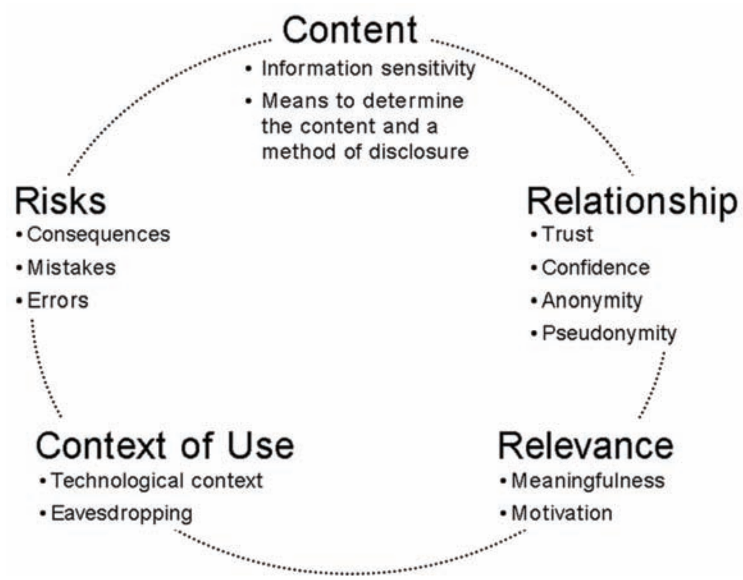
The mentioned studies and theories construct grounds for the privacy regulation model. Especially, Adams and Sasse (1999) and Adams (2000) as well as Bellotti and Sellen (1991), have pointed out essential ideas related to constructing the privacy regulation model. Further, the fundamental theoretical ideas of the following model are based on Altman's (1975) way of understanding privacy as a dialectic process through boundary regulation behavior where an individual dynamically uses various mechanisms for regulating information disclosure. The model concentrates on factors affecting output (i.e., information disclosure) and excludes input regulation (receiving information).

The privacy regulation model that have been defined consists of five basic factors that represent people's tendencies to perceive their privacy when disclosing personal information by the means of communication technology. Individuals perceive privacy and privacy regulation in their own personal way, and therefore the idea of the model is to describe factors affecting privacy regulation rather than explaining how people do that in practice. Even where the model suggests explicit factors affecting information disclosure, the process of deciding about information disclosure is often unconscious rather than conscious. In addition, the factors are partially overlapping and their priority changes from one situation to another. For these reasons, the factors are illustrated as nodes in a circle without any clear start and end points. Figure 1 illustrates the model by presenting the five factors affecting information disclosure and these are described.

Content is the information that users disclose. Examples of content include voice, text, audio, video, and images. Characteristics of content and its *sensitivity* determine people's willingness to disclose information. Overall, people are more comfortable with disclosing information that is perceived as more public than private.

Altman (1975) brought out that speech and voice intensity are ways of regulating privacy

Figure 1. The privacy regulation model



for voice content in face-to-face communication settings. Similarly, various methods of creating, editing, and selecting digital content, that will be shared or distributed to others, can be understood as privacy regulation mechanisms. For example, asynchronous text-based communication, like text messaging, leaves relatively much space for determining how individuals express themselves. In another extreme are geographical location disclosure services that detect, gather, and disseminate information automatically. In these kinds of cases, users have some or no control over determining the content that is disclosed or when disclosure takes place.

The relationship between a user disclosing information and the recipient(s) is another important factor affecting information disclosure. The most important characteristics are *trust* and *confidence*. Seligman (1997) uses the term confidence when users have strong expectations on appropriate use of the information they disclose. If users do not have expectations of any kind related to further use of disclosed information, Seligman talks about trust: "trust begins when knowledge ends," he says. Overall, the history of the relationship and the history of communication, as well as other attributes characterizing social relationships, determine individual's willingness to disclose information to the other.

According the privacy regulation model, *anonymity* and *pseudonymity* play significant roles in information disclosure. The nature of communication is remarkably different in situations where all or one of the parties are anonymous, compared to situations where true identities are used. Today, nicknames are commonly used in Web communities and leaving truly anonymous messages to discussion forums is a well-known communication practice. Anonymity sometimes makes an individual act more openly than when using a true identity. Anonym or pseudonym communication usually ignores other factors affecting information disclosure due to the fact that the content is not (easily) mapped to the user's true identity. From the privacy perspective, one of the most interesting usage behaviors is related to revealing one's true identity, like disclosing a phone number in a chat room, after using nicknames for a longer time. This requires a high level of trust (belief in the other's good purposes) and it brings the relationship between the users to a new level. There are examples of serious consequences of this kind of behavior, for example, when adults search for inappropriate relationships with minors in chat rooms and Web communities.

Currently, mobile communication solutions have not been seen to support truly anonymous or pseudonymous communication behavior, and therefore these privacy regulation mechanisms are not considered to be applicable in the mobile domain. However, the mobile Internet connection increasingly brings the social computing services to mobile phone users too, resulting the opportunity for anonymity and pseudonymity.

Relevance relates to the first two factors; *intended content* should be disclosed to *intended recipients.* The user should be able to decide what information, to what extent, and to whom it will be disclosed.

We can distinguish two main things that relevance means within the context of privacy regulation; disclosure of *meaningful information* to others and having a *motivation to disclose* information. First, the information that an individual discloses to another party needs to be relevant in the given social environment. For example, a teenager may select different images taken during a vacation to share with his grandmother and his closest friend, even if both relationships are highly trustworthy. If an individual who shares content feels that it does not belong to the other party, then privacy regulation mechanisms have failed. Second, relevance refers to motivation to disclose information, meaning that users will benefit (whether it means utility or pleasure, such as socializing) somehow from disclosing information.

A current example of privacy regulation practices addressing the issue of relevance is managing access rights levels, for example, for blogs or Internet services allowing users to upload digital images and share them with others. In the latter case, the users are able distinguish information, which is more private and meant only for selected people, and information which is more public and openly accessible to anyone.

Risk of information misuse refers to perceived possibilities to use the information for other than originally intended and harmful purposes. Risks are directly linked to the content and relationship factors. When the recipient of information is well-known, having a long history of communication and when the information is not perceived sensitive, the risks are smaller than when disclosing very private information to an unknown recipient.

Mediated communication practices also contain risks of making mistakes, such as sending an e-mail to unintended recipients, publishing one's Bluetooth identity (Persson, Blom, & Jung, 2005), or publishing images accidentally, to an image sharing service while saving the images to a private mobile blog. Therefore maintaining user's awareness of what the device is doing and what information users are publishing about themselves becomes increasingly important. Further, communication devices sometimes make errors or misinterpretations. For instance, social awareness systems may convey disinformation on user's current context, if it has misinterpreted it (see the case example of PePe presented later in this chapter).

The problems regarding recordability and unawareness of the final audience (Palen & Dourish, 2003) have direct impact on evaluating risks of disclosing personal information. This is seen in ways people regulate their privacy when using communication technologies. For instance, if a user does not want to leave a record of information s/he wishes to disclose, s/he can select a communication method that will not be recorded, like making a phone call instead of sending a text message.

The mentioned examples show that both users and designers of mobile technology are creating new control mechanisms for privacy regulation. Designers and developers have a great opportunity to create usable and acceptable means for privacy regulation, whilst the users can adapt and apply the given means to their own needs.

Context of use is a relevant element having an impact on disclosure of information. Eavesdropping (heard or seen) is a considerable factor affecting privacy perception. For instance, in a crowded place, a mobile phone user needs to asses whether the content of her speech is private and spoken only for the other party, or if it is something that can be heard by anyone in the near surroundings.

In addition, the environment may be equipped with devices and sensors (such as surveillance cameras and microphones) detecting and recording user's activities. For instance, a person entering a meeting room who does not notice that a conference call is open to a remote site may disclose information that is intended only to people sitting in the room. These kinds of failures to notice the presence of such technologies might cause some extremely inconvenient situations and even result in the invasion of privacy.

MODEL EVALUATION CASE STUDY: PRIVACY PERCEPTION OF THE PEPE SYSTEM

This section gives an example of how to apply the privacy regulation model in the evaluation of a mobile communication solution. The solution and the study are presented first in order to contextualize the evaluation of privacy perception. Then the results of the evaluation are described in detail to give an idea how to apply the model in practice. The study with specific results of the location enhanced mobile presence related user behavior is described in detail in Lehikoinen and Kaikkonen (2006).

A qualitative field study that aimed to gain understanding of the usage of a mobile social awareness system, PePe (abbreviation from the words *Presence of Peers*), was conducted in Helsinki, Finland in 2004 (see Ibid). In the study, special attention was paid to privacy regulation features and practices.

Participants

The participants were mostly high school students or recently graduated from high school. Together, they formed a coherent social network. Obviously, users formed smaller groups having closer relations with each other than with the rest of the group. Thus, existing social networks were present, which was crucial in order to gain realistic understanding of usage of the PePe system in daily life. The number of participants was 12, four were male and eight female. All users were 19 years old. Users were recruited in such a way that they naturally formed a group of friends. This ensured that the users had a large enough group of persons with whom to share sensitive information, like their current location. Young adults were recruited as participants due to the fact that social networks and social life are often important for them. At the age of 19, teenagers or young adults are no longer restricted to the small geographical area where they attend school and hobbies. Young adults have friends and acquaintances in a large area and they socialize with a large group of people. Having users that go to the same school and live in the same area would not have shown the need for viewing their location information.

The PePe System

A mobile social awareness system, PePe, ran on Symbian S60 OS mobile phones. The system allowed mobile phone users to share their status information, including location, and see others' status information. Location was detected from cell ID information that mobile phones provide without any additional devices.

The PePe system supported automatic location updates. In order to use this automatic feature, users needed to first create the locations that they intended to share with others. For instance, when a user was at school, s/he created a location called "at school" to the PePe system. Location information was then updated automatically every time, when the user arrived at school. When the user was in a location that was not created, the system considered it as 'unknown location.'

The users also had the possibility to update their status manually by the means of short status text, status image, and an availability icon. The users were able to choose between three pre-made availability icons, doors, that illustrate being available (open door), not available (closed door), and busy (half-opened door). They were also able to express their status with two lines of free form text. The status image could be taken with a camera phone or it could be ready-made graphics.

The other users were able to watch location information as well as the other presence information from a contact list on their mobile phone. Location information was shown to other users textually, exactly as the user had defined it (see Figure 2, middle).

Cell ID positioning technology was chosen because it was easily available and did not require any additinoal hardware. In a case of GPS positioning technology, a separate module would have been needed and it would have decreased the usability of the system. Furthermore, GPS did not support positioning indoors. However, the Cell ID-based locations were somewhat inaccurate. Cell ID boundaries are not clear for the users due to the fact that multiple cells sometimes overlap one location. These technological limitations caused some confusion for the users, but they did not have a remarkable impact on the results.

Privacy Regulation Features

Privacy was kept firmly in mind when the mobile presence system was designed. Fundamentally, users were able to decide whether they used the service and published their presence information or not. Further, users were able to define the information they disclosed by themselves, like naming the locations that were shared with others. The presence system provided two levels of disclosing presence information; private and public. On the private level, all presence attributes (location, availability icon, status message, and image) were disclosed. Subscription to private presence required authorization from the user. On the public level, only availability information, that is, text (available, not available, and busy) and an icon, was shared.

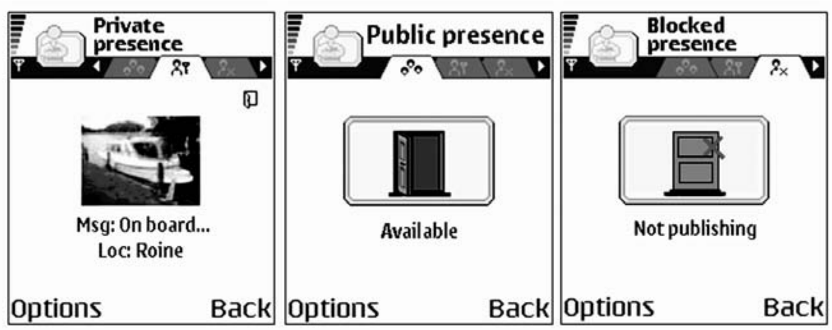
The users were also able to hide their presence information from others. In such cases, the other users were not able to see any information from the one publishing presence information. Examples of different presence statuses can be seen in Figure 2.

Procedure

The study consisted of three steps: an introduction session, a usage period of three weeks, and the final interviews. In the introduction session, the main objectives of the study and the mobile presence system were introduced to the participants. The configuration and other settings were prepared for the users so that they were able to start using the system immediately. The participants were also told how to contact the test moderator in case of a problem.

The participants were asked to use the test phones like their own phone during the pilot period and were encouraged to use the PePe system like they would use it in their everyday life. The test procedure included a task of marking the places

Figure 2a. Private presence view; Figure 2b. Public presence view; Figure 2c. Blocked presence view



(a) *(b)* *(c)*

they visited during the 3 day test period. Two of these days were predefined and the user was able to define the third one. This task made it possible to compare the locations where the participants visited in real life with those that they had defined in the PePe system.

The qualitative data was derived from one-on-one in-depth interviews. The final interviews were conducted immediately after the usage period. Each interview took from 1 to 1 ½ hours. The interview consisted of the following themes: general experiences of usage, publishing the user's own presence information, naming and using location information, watching others' presence statuses, and privacy. In the privacy theme, the privacy regulation model was utilized by asking questions addressing the factors affecting information disclosure.

The interviewees were asked to show how they used the system and give examples of both their own and other users' presence statuses. The users defined a sociogram of the relationships with the other. The interviews were recorded and transcribed.

Results: Privacy Regulation Practices in Using the PePe System

This section describes the overall results of evaluation of the privacy regulation practices regarding the PePe system, in order to give an example of utilization of the privacy regulation model (see Figure 1). All five factors are discussed separately due to clarity reasons, even if in practice they are somewhat interdependent. Further, questions addressing each of the factors are presented in the respective sections.

Content

The PePe users are able to disclose different kinds of information including availability, free edited text, image, and location information. The participants valued these differently in terms of information sensitivity. Further, each of the attributes has distinguishable mechanisms for regulating disclosure of information. The PePe attributes are analyzed separately.

The following questions were asked from the participants to address sensitivity of the content:

- How do you perceive presence information? Is it private or public?
- What is the order of each presence attribute in terms of information sensitivity? Why?

Location information was felt as the most sensitive piece of information (all of the users) due to the fact that it was also the most descriptive in terms of conveying user's context information. Location information was dynamic, and thus always up to date. By viewing it, the participants were able to determine the whereabouts of the others as well as some other contextual attributes, such as social and physical contexts. The participants defined on average, 24 locations in the PePe system, varying from 8 to 43, thus most of their frequently visited places were shared with the others.

The participants were able to select which locations they wanted to disclose with the others, and name them as they wished. Sometimes, the participants used named the locations in such a way that they were intelligible only for the other participants, who knew their living conditions and habits fairly well. In addition, the participants used very generic names for the locations they defined to the system. For instance, a location called "home," "at Krista's home," or "at work" are extremely descriptive for those who know the person, but not for anyone else. These practices were considered to be important, making location disclosure acceptable from the privacy point of view.

An important observation relates to the interpretation of "unknown" location information. Interestingly, it had a clear indication of a person's status and even whereabouts, although it did not inform anything explicitly. The fact that the participants knew each other's daily routines allowed them to infer a person's status based on an "unknown location" note. In most cases, it indicated that the user was traveling from a previous location to a new location.

To conclude, location information was considered to be highly sensitive information due to the fact that by following its changes, one was able to get a clear picture of the other user's routines. The participants regulated their privacy by choosing the locations they shared and naming them in such a way that they were not interpretable by unknown people.

Status image and **status text** were considered almost equally sensitive, but obviously the sensitivity depended on the informational content that the image or the text conveyed. Three participants used share fairly private and personal images that, for instance, described things that are important to them. The other participants shared generic and thus more public images. Also, status texts were used for different purposes, like specifying one's context for instance, by describing the activity taking place, or for greetings.

The participants were able to choose the images they would use for presence status. In all cases, the status image was taken by the camera of the phone. Similarly, the participants were able to decide what to write in the status text. Changing these attributes required manual update. The PePe system provided these mechanisms for privacy regulation and they affected positively, the participant's perception of information disclosure.

Availability information was considered to be public rather than private and all the participants said that it could be disclosed to anyone. They felt that availability is the least sensitive presence attribute. The participants changed availability information manually very rarely, and it was set as "available" by default.

Relationship

Relationships between the PePe user trial participants were analyzed by the means of a sociogram, which meant that the participants defined their relationships to the other participants. The main principles of defining the sociogram in this context were as follows; first, the closest persons were placed close to the self (center of the paper) in the sociogram and second, persons belonging to the same group in real life formed a group in the sociogram as well (to learn more about sociograms, see Moreno, 1932, 1934, and 1953). The following questions were asked of them while and after the sociogram tasks:

- How would you describe your relationship with these contacts?
- With whom did you share your presence information?
- With whom did you use PePe most often?

The sociogram tasks and the interviews brought out that the group of participants was fairly coherent (as was expected due to recruitment criteria) and each participant knew each other to some extent. However, an analysis of the sociograms showed that the group was roughly divided into two subgroups. Further, one of the participants belonged only to one participant's "inner-circle" and the rest felt her to be the most distant one. This was the framework of social relations from which privacy regulation was evaluated.

Overall, the participants described the relationships as extremely trustworthy and they were confident that the others would not misuse the information that they disclosed by means of the PePe system. Even though one of the participants was more distant than the others, she was trusted due to the fact that she was a good friend of one of the participants who happened to be a hub of the social network. Trustworthiness and confidence were reasons why all the participants let everyone view their private presence information, meaning that all participants were able to view all four attributes (location, image, text, and availability).

Relevance

Relevance of the presence information was evaluated by asking the following questions: What is the order of each presence attribute in terms of relevance? Why?

Location was mentioned by all participants to be the most relevant presence attribute (12/12 participants). It was *meaningful* for the closest friends since they were able to infer one's context based on that. The strongest *motivation* for disclosing location information was that it facilitated both intentional and ad-hoc group coordination better than the other presence attributes.

Status text and **image** came after location and before availability in the participants' list of relevant attributes. The order is difficult to state exactly, due to different habits of using the status texts and the images, as mentioned earlier. The strongest motivation for publishing a status image was to make the presence status look more personal. In contrast, status texts facilitated multiple purposes like describing current activity, conveying greetings, or informing about upcoming events.

Availability was considered to be the least relevant presence attribute. However, the participants also brought out that it is the least sensitive information, thus they did not pay any attention to its disclosure.

Sociogram revealed that PePe was utilized most with the closest friends, even though the same information was disclosed to everyone. Especially, location information including the ways to name the locations, images, and status texts were defined and written, bearing in mind how the closest friends would interpret them. This relates also to the publishing nature of information disclosure; those who are interested in the information the publisher discloses can freely go and view it, even though the same information is openly retrievable for any other user as well.

Interestingly, the order of relevance corresponds with the order of content sensitivity. This refers to strong interdependency of the content sensitivity and the relevance factors, at least when analyzing disclosure of context information.

Risk of Information Misuse

Risks of misusing the information disclosed by the means of PePe were evaluated, focusing on the participant's experiences and thoughts about the risks. This means that analysis of all the imaginable and theoretical ways of misusing the disclosed information was excluded. Risks were asked from the participants through the following questions:

- How do you assess the risks of disclosing presence information?
- How do you think that presence the information could be misused?
- Were you always aware of information you are disclosing?

The participants did not perceive major risks related to disclosing presence information. The most important factor explaining this was coherence of the group. Further, presence information was non-recordable, and thus its misuse in a form of forwarding it digitally to third parties was extremely difficult (the PePe client or the device did not provide any kinds of means for recording presence information). In addition, the participants thought they had full control over determining the information they disclosed to others.

A clear factor affecting the perceived risks of misusage of the content was being constantly aware of the information that was disclosed. Mobile presence system brings this up extremely well due to the fact that information changes according to the context. The participants were not always fully aware of the information they were disclosing and sometimes the presence status was incorrect due to inaccurate Cell ID-based location tracking. For instance, one participant's home was close to a café, where he and his friends occasionally spent free time occasionally. Sometimes the system interpreted that this participant was at the café even though he was actually at home. This naturally caused some wondering and questions about this participant's way of living. The participant did not consider that as an invasion of privacy, although it was somewhat annoying, as he pointed out. However, these kinds of problems might emerge as true privacy threats in real life usage.

Few participants brought explicitly out that they learned to know their friends better by viewing their frequently updated presence statuses. Things like, where they were spending their time, with whom they were (often friends' homes were defined as locations in the system), what they were doing, even when they went to sleep, and so forth, were revealed during the 3 weeks of usage. Although, this was a fairly generally observed fact, the participants did not consider this as a threat to their privacy.

To conclude, the participants did not perceive PePe usage as causing serious privacy risks, even though they acknowledged that the information they disclosed was fairly descriptive and personal, especially in the longer run. The strongest reasons for confidence were coherence and trustworthiness of the group and the non-recordable nature of the disclosed information.

Context of Use

Mobile awareness systems are special cases when analyzing the context of use, due to the fact that *context* is (at least partially) the *content* that is disclosed. Thus, context plays a significant role with regard to control mechanisms addressing content, but otherwise the local context of use does not greatly affect privacy perception. A reason for this is the nature of presence usage, which is not easily observable by others in close proximity, unless the users want intentionally to share what they are doing.

DISCUSSION

In many mobile communication situations, the basic privacy regulation features meet the users' needs. However, little has been done to solve the problems of recordability and unawareness of the final audience or content recipients. One of the biggest concerns in this area is publishing digital images to the Web. Today, many images are taken by camera enabled mobile phones with Internet connectivity. These make content publishing almost immediate and many questions emerge due to this, such as: Should the people who have been photographed have some means of control, or at least be aware of, further usage of the photo?; how the content will be further used; and who will see it. These problems that are not directly linked to the (first) user's privacy are still more or less unsolved.

More broadly, privacy, especially informational privacy, has already changed or diminished due to information and communication technological development. This has led a few people to ask whether we need privacy at all, and what it is needed for. A truly transparent society might even be more secure due to the fact that criminal behavior would be easier to detect and punish. However, this kind of thinking is not widely spread but it shows that technology has a clear role in shaping people's everyday life and attitudes.

Some studies have shown (e.g., Adams, 2000; Lehikoinen et. al., 2006) that users do not perceive privacy threats unless they encounter the consequence of the misuse of their personal information. This observation raises a question of how designers and developers should address *potential* privacy threats. Should they design systems that limit usage behavior in order to avoid possible invasion of privacy? If yes, then how and to what extent should usage be limited? Users of, for example, mobile communication technologies should anyhow be treated as autonomous individuals who are able to determine by themselves the level of risk they are taking in their practices. However, making the risks as visible as possible will help to avoid undesired consequences.

CONCLUDING REMARKS

This chapter extends theoretical understanding on privacy regulation in the context of mobile information and communication technology. The privacy regulation model proposed in this chapter is a combination of earlier theories presented in the fields of HCI and ubicomp and it has been extended to cover some aspects that have been missing in the previous research. The model consists of five factors: content, relationship, relevance, risks, and context. These factors build the grounds for understanding human practices related to privacy regulation in the context of, especially mobile, computer mediated communication.

The applicability of the privacy regulation model has been demonstrated in the study of privacy regulation practices of a mobile presence system. The five factors presented in the model were applied in designing the study and analyzing the data from the perspective of privacy. The privacy regulation model can also be considered to be a set of design guidelines for dealing with systems and solutions handling disclosure of personal information. Designers need to consider the five factors affecting privacy perception and apply them to both functionality and UI design of their solutions. The model facilitates designing relevant control mechanisms for information and communication technological solutions by bringing out factors affecting human behavior with regard to information disclosure.

However, privacy is an extremely complex and multifaceted concept and application of the privacy regulation model in the area of HCI design and evaluation is needed to translate the theoretical ideas to concrete privacy management features, functions, and/or designs. Further research interests are to investigate personal content disclosure by the means of mobile communication devices in the context of blogging and social computing services, and to further develop privacy regulation mechanisms that meet users' needs. The privacy regulation model provides a useful basis for such further research.

REFERENCES

Adams, A. (1999). The implications of users' privacy perception on communication and information privacy policies. In *Proceedings of the Telecommunications Policy Research Conference*.

Adams, A. (2000). Multimedia information changes the whole privacy ballgame. In *Proceedings of the Tenth Conference on Computers, Freedom and Privacy* (pp. 25-32).

Adams, A., & Sasse, M. A. (1999). Taming the wolf in sheep's clothing: Privacy in multimedia communications. In *Proceedings of ACM Multimedia* (pp. 101-107).

Al-Muhtadi, J., Campbell, R., Kapadia, A., Mickunas, M. D., & Seung Yi. (2002). Routing through the mist: Privacy preserving communication in ubiquitous computing environments. In *Proceedings of the International Conference of Distributed Computing Systems* (pp. 65-74).

Altman, I. (1975). *The environment and social behavior.* Monterey, CA: Brooks/Cole Publishing.

Beckwith, R. (2003). Designing for ubiquity: The perception of privacy. *Pervasive Computing, 2*(2), 40-46.

Bellotti, V., & Sellen A. (1993). Design for privacy in ubiquitous computing environments. In *Proceedings of the ECSCW* (pp. 77-92).

Hong, J. I., Ng, J. D., & Lederer, S. (2004). Privacy risk models for designing privacy-sensitive ubiquitous computing systems. In *Proceedings of Designing Interactive Systems* (pp. 91-100).

Jiang, X., Hong, J. I., & Landay, J.A. (2002). Approximate information flows: Socially-based modeling of privacy in ubiquitous computing. In *Proceedings of UbiComp 2002* (pp. 176-1993).

Langheinrich, M. (2002). A privacy awareness system for ubiquitous computing environments. In *Proceedings of UbiComp* (pp. 237-245).

Lederer, S., Dey, A. K., & Mankoff, J. (2002). *Everyday privacy in ubiquitous computing environments.* Paper presented at the Workshop on socially-informed design of privacy-enhancing solutions in Conference on Ubiquitous Computing. Retrieved October 9, 2006, from http://guir.berkeley.edu/pubs/ubicomp2002/privacyworkshop/

Lehikoinen, J. T., & Kaikkonen, A. (2006). PePe field study: Constructing meanings for locations in the context of mobile presence. In *Proceedings of the 8th Conference on Human-computer interaction with mobile devices and services* (pp. 53-60).

Lessig, L. (1998). The architecture of privacy. In *Proceedings of Taiwan NET.* Retrieved January 3, 2007, from www.lessig.org/content/ articles/works/architecture_priv.pdf

Marx, G. T. (2001). Murky conceptual waters: the public and the private. *Ethics and Information Technology*, *3*(3), 157-169.

Moreno, J. L. (1932). *Application of the group method to classification*. New York: National Committee on Prisons and Prison Labor.

Moreno, J. L. (1943). *Who shall survive?* Washington, DC: Nervous and Mental Disease Publishing Company.

Moreno, J. L. (1953). *Who shall survive?* Beacon, NY: Beacon House Inc.

Palen, L., & Dourish, P. (2003). Unpacking "privacy" for a networked world. In *Proceedings of the ACM CHI Conference on Human Factors in Computing Systems* (pp. 129-136).

Persson, P., Blom, J., & Jung, Y. (2005). DigiDress: A field trial of an expressive social proximity application. In Beigl et al. (Ed.), *UbiComp 2005: Ubiquitous computing, LNCS*, 3660, 195-212.

Price, B., Adam, K., & Nuseibeh, B. (2005). Keeping ubiquitous computing to yourself: A practical model for user control of privacy. *International Journal on Human Computer Studies*, 228-253.

Schoeman, F. D**.** (1992). *Privacy and social freedom*. Cambridge: Cambridge University Press.

Seligman, A. (1997). *The problem of trust*. Princeton: Princeton University Press.

Soppera, A., & Burbridge, T. (2004). Maintaining privacy in pervasive computing—Enabling acceptance of sensor-based services. *BT Technology Journal*, *22*(3), 106-118.

Westin, A. (1970). *Privacy and freedom* (2nd ed.). New York: Atheneum.

KEY TERMS

Control Mechanism: A method that an individual uses to determine the level of social interaction with others. For example, verbal behavior in face-to-face context and access rights managements in the context of computer mediated communication control mechanisms.

Design Problem: An unsolved state or an issue that a system being designed needs to take into consideration.

Mobile Presence: A service that allows users to share their context information with other users. The context information may be automatically detected (e.g., location) or manually defined (e.g., text and image).

Privacy: One of the most frequently used definitions of privacy by Alan Westin (1970) states: "Privacy is the claim of individuals to determine for themselves when, how and to what extent information about themselves is communicated to others." This definition is applicable to many information technological contexts. One of the key elements of privacy is its dialectic nature.

Privacy Regulation: The act of determining and modifying what kind of interaction and to what extent an individual interacts with others, that is, usage of a set of control mechanisms.

Privacy Features: Methods that are designed and implemented for users of computational systems to regulate their privacy in computer mediated communication.

Qualitative Method: A research method that emphasizes meaning, quality, and context. A qualitative study can consist of, for example, interviews and observations. For instance, ethnographical studies consist of a set of different qualitative methods. Qualitative data can be, for instance, words, images, impressions, gestures, and tones (e.g., voice intensity).

Chapter LII
Framework and Model of Usability Factors of Mobile Phones

Dong-Han Ham
Middlesex University, UK

Jeongyun Heo
MC R&D Center LG Electronics, Korea

Peter Fossick
Middlesex University, UK

William Wong
Middlesex University, UK

Sanghyun Park
MC R&D Center LG Electronics, Korea

Chiwon Song
MC R&D Center LG Electronics, Korea

Mike Bradley
Middlesex University, UK

ABSTRACT

This chapter aims at developing a framework and model for identifying and organizing usability factors of mobile phones. Although some studies have been made on evaluating the factors, there is no systematic framework for identifying and categorizing them. This chapter proposes a conceptual framework which has multiple views to explain different aspects of the interaction between users and mobile phones, and *which describes the world of usability factors based on these views. The multiple views include user view, product view, interaction view, dynamic view, and execution view. Furthermore, based on the conceptual framework, a multi-level hierarchical model which classified usability factors in terms of goal-means relationships was developed. Next, two case studies are described, where the usefulness of the framework and model could be confirmed. Lastly, a set of checklists which make the framework and model more practical were developed.*

INTRODUCTION

Usability has been regarded as one of the most important attributes affecting the quality of mobile phones and thus users' satisfaction (Ketola & Röykkee, 2001; Lindholm et al., 2003). There is no universal definition of usability, but the usability concept specified in ISO/IEC 9126 (1998) is now widely accepted (Schoeffel, 2003). ISO/IEC 9126 defines usability as 'the capability of the software product to be understood, learned, used and be attractive to the user, when used under specified conditions.' Although it is the definition focusing on a software product, the definition can be applied to mobile phones taking into consideration features specific to mobile phones.

Like other quality attributes, we can view usability both from design and evaluation perspectives (Folmer et al., 2003). Usability is one of a range of non-functional requirements, such as safety and security, which should be satisfied as part of the design process. Therefore, it should be properly specified during requirements analysis and designed during the architectural and implementation phases. Conversely, usability needs to be evaluated from a user-centric point of view during all the phases of design life cycle. User perception of usability is influenced by many design factors including visual appeal, hedonic qualities, logical task sequences, and pleasure in use, as well as contextual factors including the users' environment. Thus, it can be said that usability is not an absolute concept determined by the design activity only, but rather the relative concept that can be affected by unspecified factors.

To enable more systematic usability evaluation, a lot of studies examined factors or dimensions constituting usability and their relationships (Bevan, 1999). For example, ISO/IEC 9241 (1998) defines three dimensions: effectiveness, efficiency, and satisfaction. Nielsen (1993) gives another example of such factors: learnability, efficiency of use, memorability, errors, and satisfaction. These dimensions can be categorized into either objective or subjective dimension. An objective dimension generally aims to evaluate how well users conduct their tasks with the use of performance measures like task completion time and the number of errors. However, objective dimensions do not always predict the user's assessment of usability because it does not reflect users' feeling or satisfaction. Subjective dimensions therefore, need to be assessed to provide a holistic and complete usability evaluation (Treu, 1994).

Usability evaluation methods can be classified into three types: usability testing, usability inquiry, and usability inspection (Zhang, 2003). Usability testing employs representative users on typical tasks using a system or a prototype and then evaluates how user interface supports the users to do their tasks. Typical methods include co-discovery learning, question-asking protocol, and shadowing method. Usability inquiry talks to users, observes their using a system in real work, and lets them answer questions in order to understand users' feelings about the system and their information needs. Field observation, focus groups, and questionnaire survey are typical usability inquiry methods. In usability inspection, usability experts examine usability-related aspects. Typical methods are cognitive walkthrough and heuristic evaluation. It cannot be said that one method is the best in all situations. It is thus necessary to choose an appropriate method, taking into account evaluation purposes, available time, measures to be collected, and so on.

So far the general usability concepts are described without considering the features peculiar to mobile phones. To examine the usability of mobile phones, the user interface of mobile phones, functions or tasks, and the context of use of mobile phones needs to be understood.

Mobile Phones and Tasks

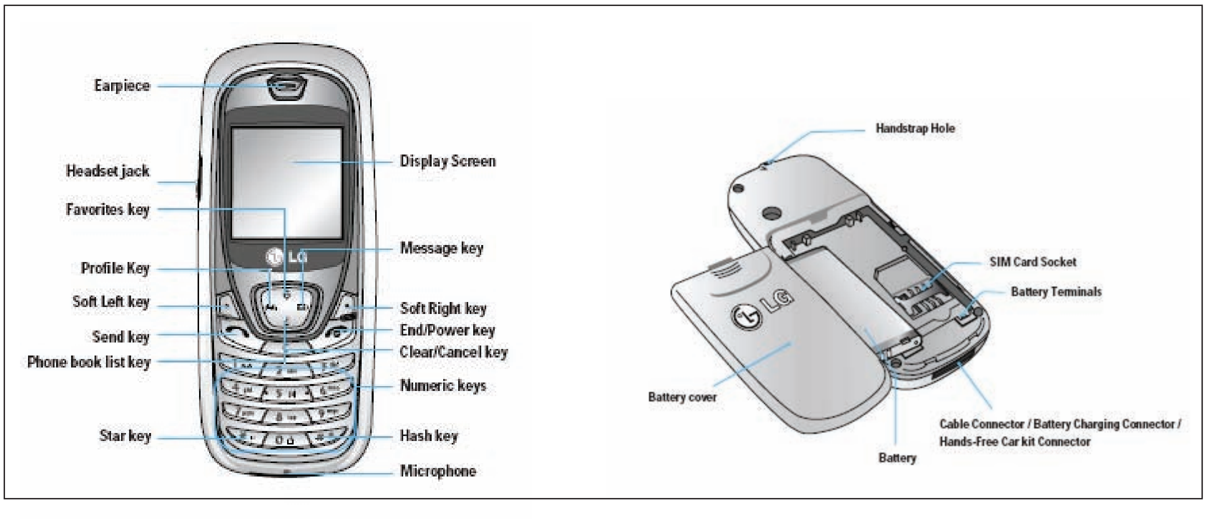
Mobile phones are portable, self-contained information and communication systems. They are characterized by three features: (1) they are used primarily in a user's hands, (2) they are operated without cables, and (3) they support the addition of new application and Internet connection (Weiss, 2002).

User Interface of Mobile Phones

Figure 1 shows a typical user interface of a mobile phone (LG-U2050 User Guide, 2005). Ketola and Röykkee (2001) divided user interface elements into seven categories from the viewpoint of their function: input (e.g., softkey, alphanumeric keys, and navigation tools), display (e.g., icons and indicators), audio and voices (e.g., ringing tones and microphone), ergonomics (e.g., touch and feeling, and slide), detachable parts (e.g., SIM card and battery), communication method (e.g., Bluetooth), and applications (e.g., making a call, games). They pointed out that user interface is just one of the interaction elements affecting usability. Other interaction elements include external interface and service interface. External interface contains user support, accessories, and supporting software. Service interface refers to service provider's services.

The seven categories described are certainly useful to understand the functional characteristics of mobile phone user interface. However, the categories can be generalized into the three user interface types that are more helpful to study users' interaction with mobile phones (Kiljander, 2004). The three types are logical user interface (LUI), graphical user interface (GUI), and physical user interface (PUI). In this study, LUI (e.g., menu and navigation structure) is defined as interface related to information contents and structure for task execution. GUI (e.g., icon and font) is defined as interface concerned with graphical or visual items presenting task-relevant information. Lastly, PUI (e.g., keypad and microphone) means tangible, physical interface properties supporting users' physical operation needed for carrying out tasks. One interface element can have properties related to two or three interface types at the same time. For example, menu is one of the LUI and has some PUI elements like icons, fonts, and colour as well.

Figure 1. User interface of mobile phones



There are three features or constraints of user interface which can affect the usability of mobile phones (Lee et al., 2006). The first is that mobile phones have too little screen to display a lot of information at the same time; therefore, information organization and navigation are critical usability issues. The second is that a physical button or key has generally more than one control function. Thus the functions of a single key are dependent on types of modes. The third is the limit in processing power and available memory.

Interacting with Mobile Phones

The increasing number of functions of mobile phones enables us to do various tasks. However, typical functions of mobile phones include making a call, sending a message, managing personal information, listening music, setting a phone configuration, taking a picture, and playing a game. Other advanced functions are watching TV, remote controlling of home automation, Internet banking, personal computing, and so on. In the forthcoming ubiquitous computing environment together with the continuing pattern of convergence, mobile phones are expected to play a key role accessing a wide range of services and will thus be an essential device of daily life.

The usability of mobile phones should be evaluated with proper consideration of context of use. Context of use means an actual condition under which a specified mobile phone is used (Dillon, 2001). It can usually be described by '5W1H' (Who, What, When, Why, Where, and How). Referring to the terminology from the area of software quality, context of use is highly related to the concept of quality-in-use rather than external or internal quality (Bevan, 1999). Context of use is constructed by various different kinds of factors (e.g., user interface elements, task types, user groups and their preferences, social constraints). However, our belief is that user interface elements and task types should be a central concept of context of use.

Research Problems

Many studies have examined various kinds of factors characterizing the usability of software or IT artefacts (Frøkær et al., 2000; Klocktar et al., 2003; Bosch & Folmer, 2004; Hornbæk, 2006). In this study, the factors are called 'usability factors.' Typical instances of the usability factors can be found in heuristics or questionnaires, which have been developed for testing software user interface design (Perlman, 1998). As an example, PHUE (Practical Heuristics for Usability Evaluation) specify 13 factors, such as help and documentation, simple and natural dialogue, intuitive mappings, and minimal memory load (Perlman, 1994). It should be noted that most of them are a type of design principle or rule. This means that they do not give much consideration to features or functions which are another aspect affecting the usability.

Such a problem can also be found in the studies dealing with the usability factors of mobile phones. Ketola and Röykkee (2001) proposed seven usability factors of mobile phones: integration of functionality, availability, utility and ease of using services, readiness for use, informativity, usefulness of support material, and interoperability. In contrast, some studies considered the usability factors by focusing on the features or functions of mobile phones without clear connection to the conceptual abstract usability factors at the level of design principles (e.g., Klockar et al., 2003).

To resolve this problem of bridging conceptual usability factors and feature-related factors, several studies proposed a hierarchical model of usability factors (ISO/IEC 9126-1, 1998; Keinonen et al., 1998; Doyanee et al., 2002; Folmer et al., 2003). For example, Doyanee et al. (2002) developed QUIM (Quality in Use Integrated Measurement), which is a hierarchical model for measuring usability of interactive software. The top level has 10 factors, which are not directly measurable, such as portability, efficiency, satisfaction, learnability, and accessibility. The 10 factors are decomposed into a total of 26 measurable sub-factors, which are further broken down into 127 specific metrics.

Previous studies on usability factors contributed towards understanding what could constitute the usability factors; however, very few attempts have been made at developing a comprehensive, conceptual framework for developing models of the usability factors, which usability engineers actually need. Additionally, most of those studies have focused on software products rather than mobile phones. Although they are helpful to identify and classify usability factors of mobile phones, it seems to fail to reflect mobile phone characteristics comprehensively.

This study firstly proposes a conceptual framework, from which several models of usability factors can be derived. And then a hierarchical model of usability factors is explained, which connects perceived usability to physically designed features of mobile phones and an extended model, which addresses the dynamic view of mobile phone use. Lastly, two case studies that were conducted to verify and improve the proposed framework are presented.

DEVELOPING A CONCEPTUAL FRAMEWORK

A conceptual framework is different from a methodological framework in that it prescribes features or requirements to be represented in a model. In

Table 1. Axioms and our approach

Axioms (& references)	Approach
Usability is emergent property (Lynch & Gillmore, 2006; Lewis, 2002)	Usability cannot be absolutely evaluated. Instead, several evaluation results based on *several viewpoints* can indicate it
Perceived usability has abstraction hierarchy characterized by goal-means relationships (Donyaee et al., 2002; Roth et al., 2001)	*User view* reflects this axiom
Usability is affected by all design-related features of mobile phones (Schoeffel, 2003; Han et al., 2001)	*Product view* reflects this axiom
Functionality restricts usability and is one critical factor of usability as well (McNamara & Kirakowski, 2006; Bawa, 2006)	*Interaction (task) view* reflects this axiom
Degree of usability changes along time (Fabre, 2000; Jordan, 2000; Hornbæk, 2006)	*Dynamic view* reflects this axiom
Usability is dependent on user groups (Dillon & Watson, 1996; ISO/IEC 9126-4, 2001)	*User and Dynamic view* reflects this axiom
Usability impact factors have a different meaning to users in terms of their preference or importance (Keinonen, 1998; Ling et al., 2006)	*Execution view* reflects this axiom (quantification scheme)
Designed usability is the outcome of constrained optimization (Folmer et al., 2003)	*Execution view* reflects this axiom (weighting value consideration)
Good usability means that perceived usability is greater than expected usability (ISO/IEC 9126-4, 2001; Hornbæk, 2006)	*Execution view* reflects this axiom (a sort of checklist)
Most of critical usability problems can be found in analytical evaluation without actual observation (Lauesen, 2005)	*Execution view* reflects this axiom (a sort of checklist)
Usability has a meaning only under specified context of use (Blom et al., 2005; Bevan, 1999)	Evaluation results using a set of checklist should be interpreted with context of use given

contrast, a methodological framework prescribes a process or procedure to develop a model (Vicente, 1999). In this regard, the proposed conceptual framework is just intended to specify information requirements that a usability factor model of mobile phones should represent; it is not concerned with detailed processes to develop a model.

A conceptual framework has been developed, taking account of two points. The first consideration is that the conceptual framework should be based on axioms about the usability of mobile phones. In this study, the axioms mean a set of facts that can be assumed from the previous studies on this topic, experience, and opinions from usability engineers in mobile phone companies. Table 1 summarizes the axioms, their related references, and the approach to dealing with them in this study. The second point is that the conceptual framework should reflect several views, each of which mirrors the different aspects of mobile phone usability. The idea underlying the second point is similar to '4+1 view' of software architecture (Albin, 2003). The views are included as the parts of the axioms, as shown in Table 1.

Conceptual Framework

Figure 2a shows the conceptual framework that is composed of five views. Four views (user, product, interaction, and dynamic) represent certain aspects of interaction between users and mobile phones, whereas the execution view reflects the actual use of the framework. As stated previously, it is intended that the framework could be a basis for developing various usability factor models.

User view reflects how users perceive the usability of mobile phones, so it interprets the usability from the user point of view, such as cognition and emotion. Product view focuses on the usability exhibited by the mobile phone itself. Thus, this view is related to technical design knowledge and areas of mobile phones. Interaction view regards tasks as a core of usability and addresses how task characteristics can affect the usability. Usability can change along usage time phase and according to user groups. Dynamic view deals with such a dynamic nature of usability. Lastly, execution view provides methods or techniques that apply usability factors to evaluate the usability of mobile phones. One fundamental idea of this framework is that any combination of five different views can result in a meaningful model of usability factors.

Figure 2b illustrates how to use different views to derive a model of usability factors. Firstly, usability factors related to each view need to be identified. Secondly, these factors can result in various models of usability factors according to their combination or priority of specific views. This study developed two models: a hierarchical model

Figure 2a. Conceptual framework— Five views of conceptual framework

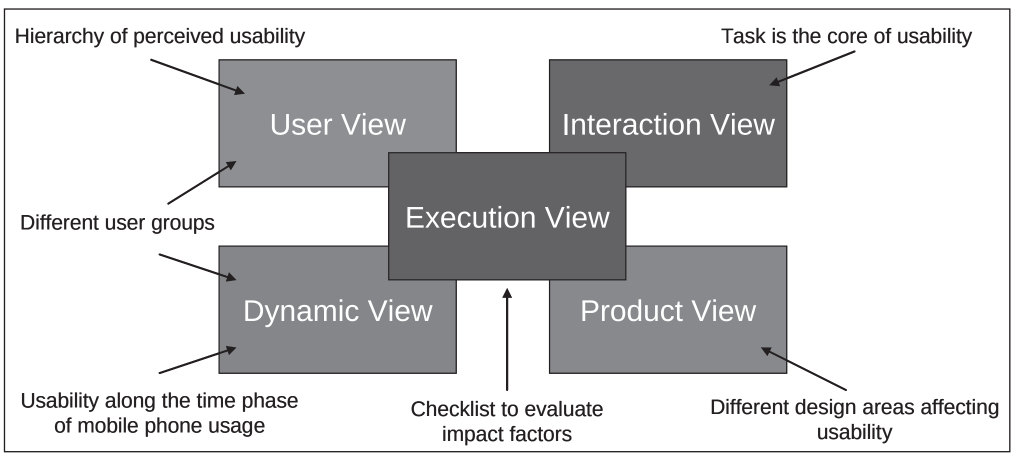
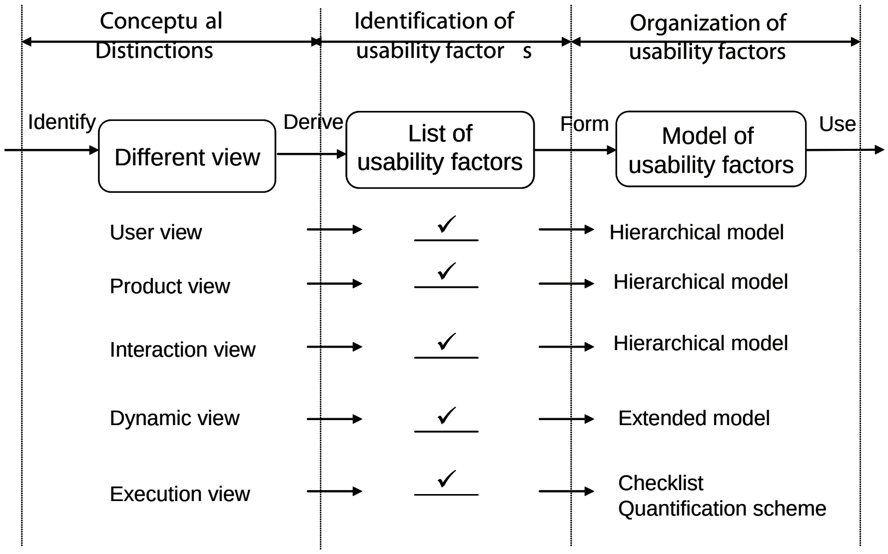


Figure 2b. Conceptual framework: Use of conceptual framework



and an extended model. The hierarchical model uses and relates three views: user view, product view, and interaction view, whereas the extended model additionally considers a dynamic view to the hierarchical model. It should be noted again that the proposed models are only examples of what could be derived from the conceptual framework. Several models with emphasis on specific views or their relationships can be developed.

HIERARCHICAL MODEL OF IMPACT FACTORS

This section explains how the conceptual framework to develop the proposed hierarchical model was used, what the main characteristics of the model are, and how it can be used for usability evaluation.

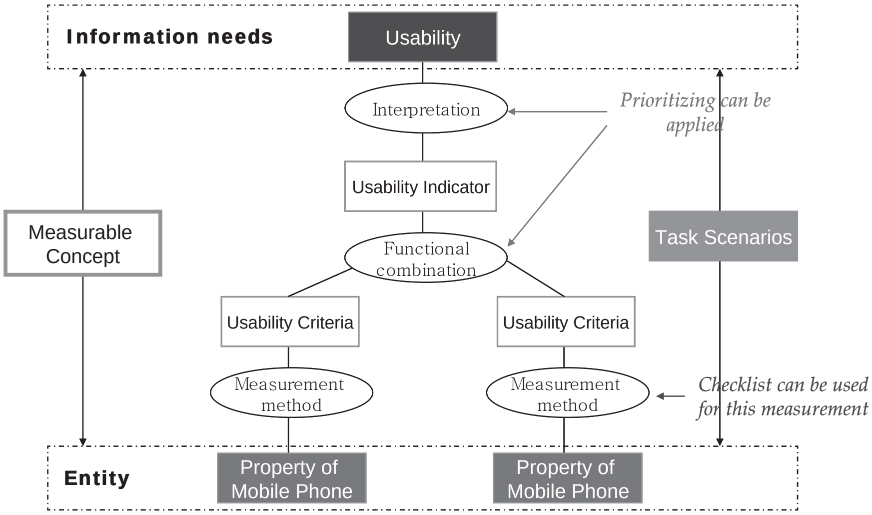
Linking Different Groups of Usability Factors

Literature review on the usability factors led to the conclusion that there is a difference in the abstraction level between usability factors. Identifying the abstraction level of each factor and organizing the factors according to the abstraction level could be a useful approach to linking different groups of usability factors. However, such organization of the factors in terms of the abstraction level is not much concerned with dynamic and execution views. Thus, only three views were used for this purpose.

Three views of the framework indicate that usability factors can be categorized into three groups. The first group is human perceived usability (user view); of which typical examples include effectiveness, efficiency, and memorability. The second is property exhibited by mobile phones (product view), and examples are reliability, durability, performance, and aesthetics. The third is performance on the tasks (interaction view), and examples are task supportability and error prevention.

Here, the arising problem is how to link these different usability factor groups, that is, how to evaluate perceived usability by measuring property exhibited by mobile phones and referring to the performance of tasks. Figure 3 illustrates the approach to addressing this problem. The approach is hinted by software measurement process framework (ISO/IEC 15939, 2001) and abstraction

Figure 3. How to link different impact factor groups



hierarchy (AH), which is a popular concept in the area of cognitive systems engineering.

In Figure 3, the final information product that needs to be known is the usability of a mobile phone and the object that needs to be examined is the usability properties of a mobile phone. The semantic difference or abstract relationship between these two concepts (i.e., usability and usability properties) are called measurable concept. Bridging the two concepts can be achieved by employing usability criteria, usability indicator, and task scenarios. Usability can be interpreted by integrating usability indicators. Usability indicators can be obtained by the functional combination of usability criteria. Usability criteria can be measured by applying measurement methods to the properties of mobile phones. These properties have specific value as usability data. From the description, five abstraction levels of usability factors can be built, where the highest is usability and the lowest is usability data. Figure 4 explains five abstraction levels of usability factors.

Here, the important point to note is that the five abstraction levels have many-to-many goal-means hierarchical relationships which are similar to AH. AH is a multilevel knowledge representation framework for describing the functional structure of a particular work domain or system (Rasmussen, 1985). One particular feature of AH is that it is defined by goal-means relationships between adjacent levels, with higher levels containing purpose-related functional information and lower levels representing more physically implemented information. AH concept has much been applied to complex socio-technical systems. Typically, five abstraction levels are known to be useful for describing the complex systems as follows:

- **Functional purpose:** The purpose for which a system is designed
- **Abstract function:** The causal structure of the process in terms of mass, energy, information, or value flows; the values or priorities that must be preserved in carrying out the work of a system
- **General function:** The purpose-related functions that the system is designed to achieve
- **Physical function:** The characteristics of the components and their interconnections
- **Physical form:** The appearance and spatial location of those components

The abstract function is different from the general function in that it is independent of a particular family of systems or products. Thus, the abstract

Figure 4. Five levels of hierarchy model

Usability (Quality in Use)	Emergent feature to be characterized with several context factors
Usability Indicator	Hypothetical, abstract conceptual constructs that are not directly measured
Usability Criteria	Sub indicator that can be directly measured through at least one specific usability property
Usability Property (Metric)	Metrics that can be obtained by observing usability data or using a formula that defines a function
Usability Data	Usability Variable that can be obtained from mobile phones, user manual, user tasks, and so on

function does not address functional characteristics limited to mobile phones. It deals with functions or quality features that can be found in any kind of information appliances. The distinction between the general function and the physical function is that the former is about a family of systems or domains and the latter is about a particular product within systems or domains. The general function is thus concerned with functions or features pertaining to mobile phones, irrespective of a particular brand or model. But the physical function is related to a specific brand or model.

To apply the AH concept in this study, we regarded the world of usability factors as a system. Although it is not a rule to consider five abstraction levels for all kinds of systems, five abstraction levels of usability factors were built. If the meaning of each abstraction level of usability factors is considered, it can be said that each of them is roughly identical with each level of AH. Thus, we can have the following mapping relationships: usability–functional purpose, usability indicator-abstract function, usability criteria-general function, usability property-physical function, and usability data-physical form. Three abstraction levels, which are located in the middle of the hierarchical model, are particularly important to evaluate usability. The three levels are described in more detail.

Usability Indicator

While usability cannot be accurately and fully evaluated in any way, it can be estimated by some usability indicators which provide a basis for decision making. Having the characteristics corresponding to the abstract function of AH, a usability indicator is not limited to the usability features exhibited by mobile phones. It is concerned with usability factors explaining how human users perceive or experience the usability of IT artefacts without being affected by peculiarities of a particular product family (mobile phones in this study). It can therefore be said that usability factors at the level of a usability indicator are applied to any other products or systems such as a personal digital assistant. On the basis of the literature review results, it is assumed that the usability indicator could be best explained with five usability factors (but this is not definitive), which are further divided into two dimensions. The five indicators are: effectiveness, efficiency, learnability, satisfaction, and customization. Of those, the first three indicators are related to task performance dimensions and they are easy to quantify. But the latter two indicators addressing emotional human factors are not easy to quantify (Keinonen, 1998; Hornbæk, 2006).

It should be again noted that the usability indicator lies at the abstraction level that does

Figure 5. Design areas of mobile phones

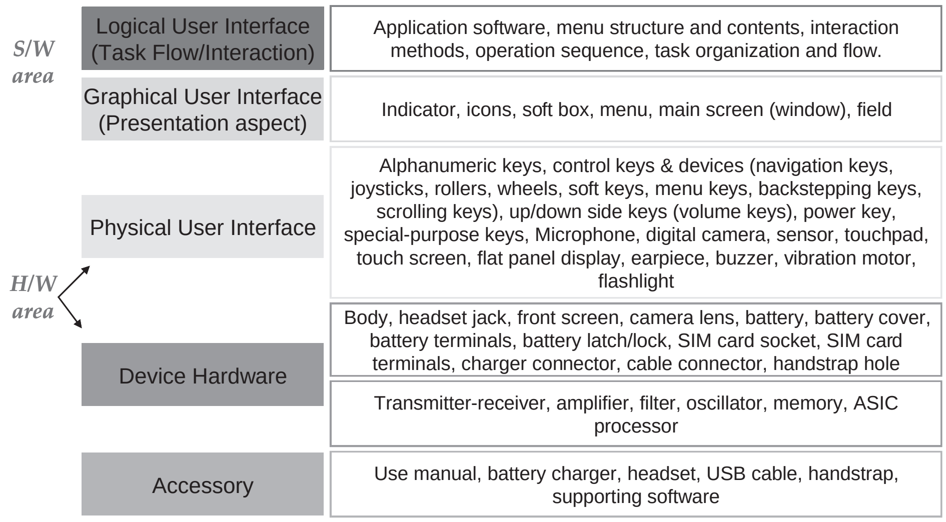
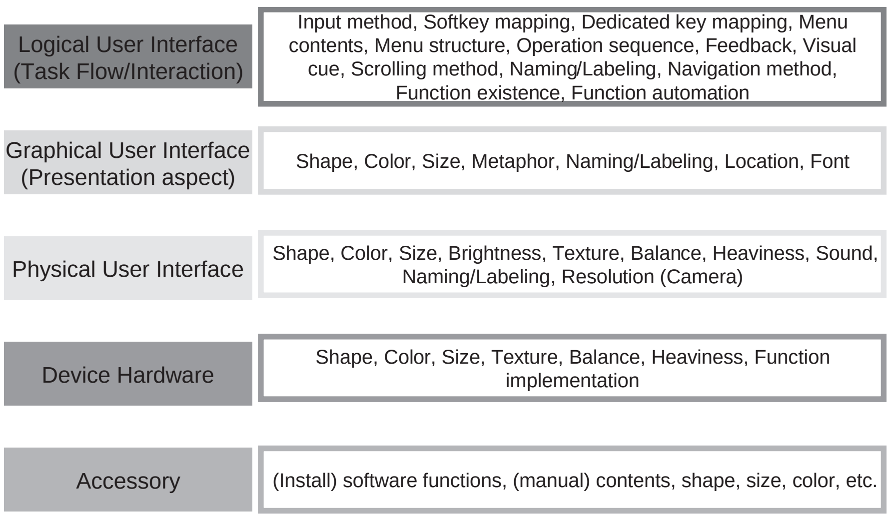


Figure 6. Factors at usability property level



not depend on the kinds or features of a product family. Thus, it serves to bridge the gap between perceived usability and usability dimensions dependent on actual products. In contrast, usability properties and criteria are dependent on actual products. While the usability criteria are the abstraction level across all kinds of mobile phones, usability properties are directly related to a specific mobile phone product. Thus, usability properties and usability criteria connect general characteristics of mobile phones and particular features of a mobile phone product. For this reason, more explanation about various design areas of mobile phones is needed to understand usability properties and usability criteria. Figure 5 shows detailed design areas.

Usability Property

Usability factors at the level of usability property measure observable features of a particular mobile phone model. They can be obtained by observing mobile phones or additionally applying some measuring functions (see Figure 6). Usability data below the level of usability property provide actual value for usability property. For example, one usability factor at the level of usability property is colour, and usability data for this factor can be blue, red, and so on. One thing to note is that usability property can be measured without carrying out tasks. This is another point to distinguish usability property from usability criteria.

Usability Criteria

As shown in Figure 7, usability factors at this level evaluate how well usability property is designed to enhance the usability of a mobile phone in consideration of task scenarios. Thus, usability criteria are aimed to offer information that is useful to assess how actual observed features of mobile phones contribute to usability indicator. For example, 'minimalism' can be measured, which can highly influence 'effectiveness' and 'efficiency' by considering softkey mapping, menu contents, and scrolling method.

Overall Hierarchical Model

Figure 8 shows the overall hierarchical model. To emphasize again, the main characteristics of the hierarchical model are the many-to-many goal-means relationships between adjacent levels and the distinction of design areas. Thus, the model seeks to elucidate how the invisible relationships between different user interface types or design areas affect the usability of mobile phones.

The proposed model is similar to the models proposed by Doyanee et al. (2002) and Bosch and Folmer (2004). But the main difference between those models and this model is that they focused on software, not mobile phones. Therefore, they do not reflect the features specific to mobile phones. Another difference lies in that they do not have explicit goal-means relationships. Lastly, they were not developed analytically from a comprehensive conceptual framework like the one developed in this study.

Use of Hierarchical Model

The proposed hierarchical model can be used as an analytical tool for several usability evaluation activities. Here two possible uses are described: interpreting evaluation items and analysing usability problems.

Figure 7. Factors at usability criteria level

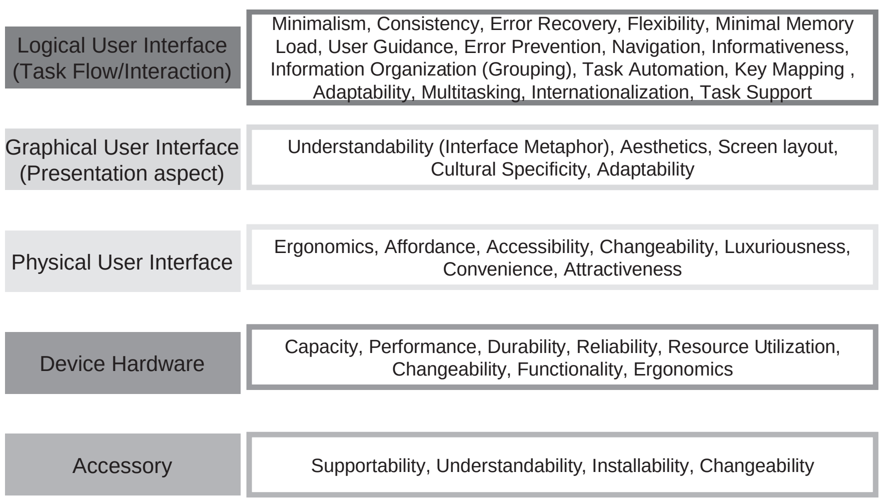
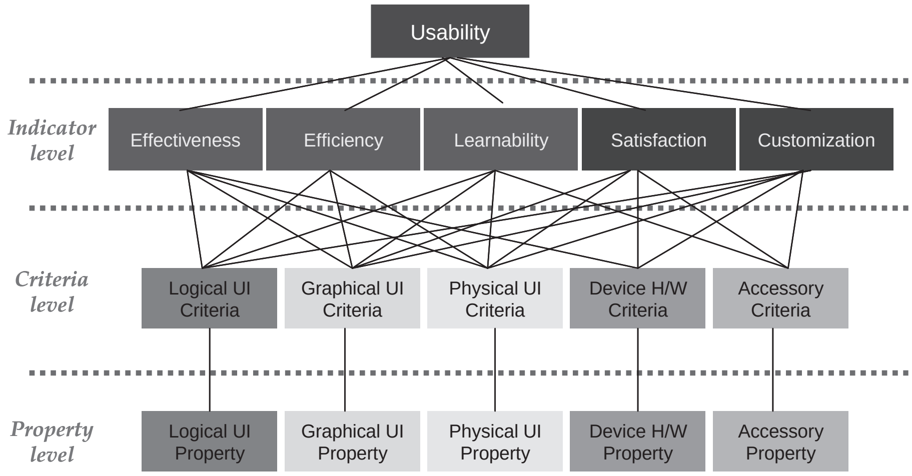


Figure 8. Hierarchical model of usability impact factors



Interpreting Evaluation Items

The hierarchical model can be used for the purpose of interpreting usability evaluation items in a more systematic way. For example, as shown in the left of Figure 9, let's suppose the case where it needs to be examined how "functional support of Bluetooth" affects the usability of a mobile phone. Although several factors at the level of usability property can be associated with "functional support of Bluetooth," two most plausible factors based on design knowledge of mobile phones, which are "functional existence" of LUI and "function implementation" of device hardware can be chosen. These two factors are then connected to "task support" of LUI and "functionality" of device hardware, at the level of usability criteria in the hierarchy, and they have some impact on "effectiveness" at the level of the usability indicator. In this way, the hierarchical model helps to understand what mobile phone features are related to an evaluation item and how the effects of an evaluation item propagate to the usability. The example on the right of Figure 9 can be explained likewise. Thus, it can be said that a usability factor related to the evaluation item "sound quality of MP3" at the level of the usability property is "function implementation." This factor influences the "performance" and further, on the "satisfaction" factor.

Analysing Usability Problems

Another use of the hierarchical model is to analyse the causes of a usability problem and the effects of the problem on the usability. Figure 10 shows two examples. The first usability problem is that "the soft key menu 'options' varies depending on the file view." The possible causes of this problem at the level of usability property include "dedicated key mapping," "navigation method," and "naming/labelling." These causes can have an influence on "navigation" and "user guidance," and further, on "effectiveness." The causes of the second problem in Figure 10 can be analysed in the same way. Sometimes users were often unsure which button to choose for going back to the previous page or for closing the running task. The causes at the level of the usability property can be "soft key mapping" and "menu structure." They affect "navigation" at the level of usability criteria and further, "effectiveness." Although an exact cause of a usability problem cannot be found by using the hierarchical model, it can be a useful tool for analysing the problem. One thing to note is that determining goal-means relationships between usability factors are dependent on analysers' knowledge on mobile phones and tasks.

Figure 9. Interpreting evaluation items

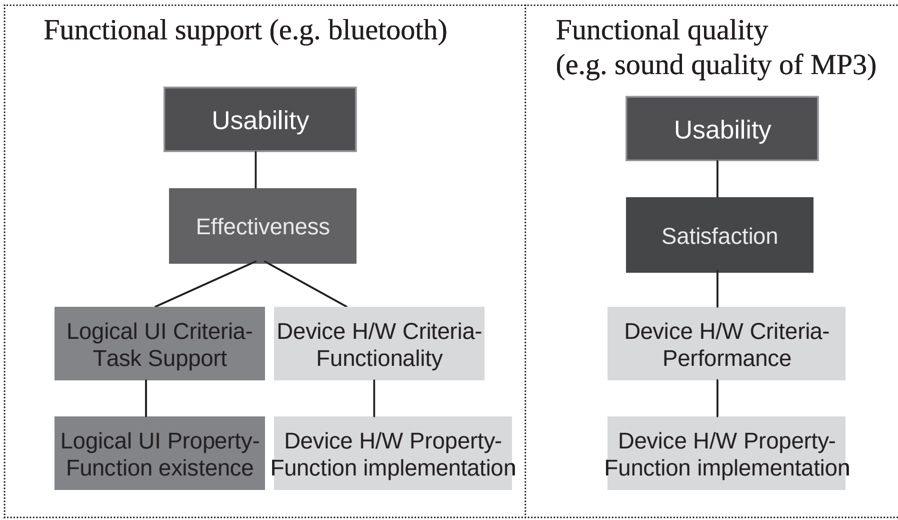
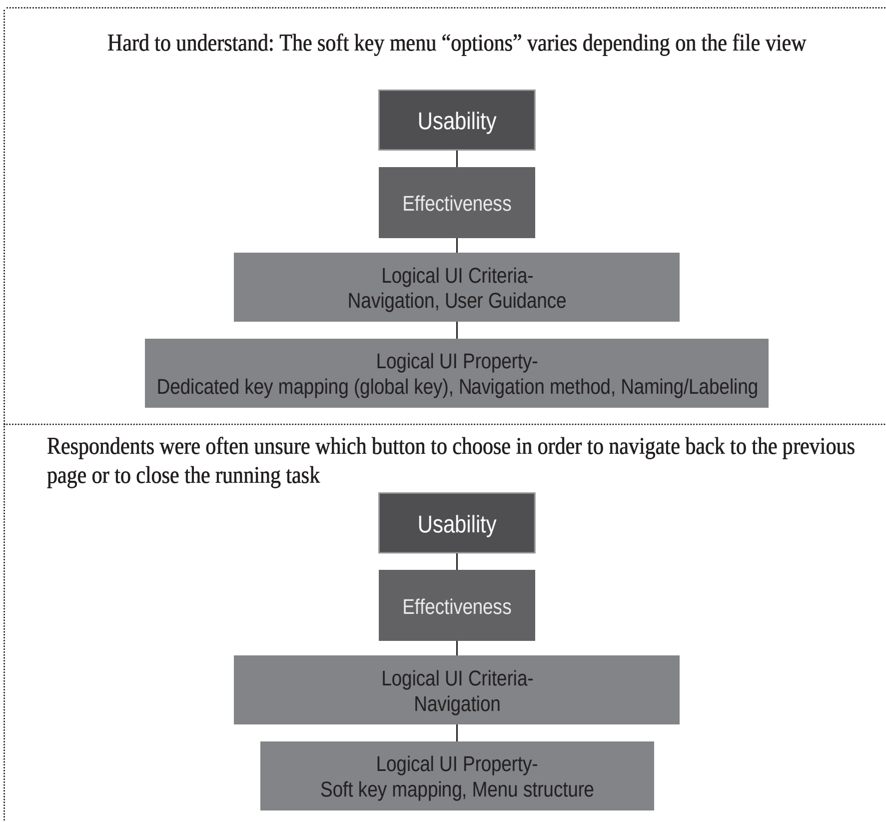


Figure 10. Analysing usability problems



CASE STUDIES

The 1st Case Study

The main purpose of this case study was to assess the practicality of the proposed conceptual framework and the hierarchical model. In this case study, 21 undergraduate students (male: 16, female: 5) majoring in product design at Middlesex University, London, in the United Kingdom, served as participants. Their age ranged from 19 to 25 excepting one person, and the usage period of their own mobile phone ranged mainly between 6 and 12 months. Of those, 12 persons used Nokia phones, two used Samsung, three used Sony-Ericsson, three used Motorola, and one used Siemens.

Two participants were grouped into one team, and one participant conducted a task while the other observed it. After one participant finished the task, the roles were then reversed. This was repeated for seven tasks and their actions were recorded using a video camera. The seven experimental tasks were: making a call, adding a contact to an address book, inputting a text message and sending the text, using PIM (Personal Information Management), using WAP (Wireless Application Protocol) services, taking a picture, and changing a phone setting. For the experiment, the participants used their own mobile phones. It therefore should be noted that the data collected for each task is different because tested mobile phones did not provide the same user interfaces and set of functions. However, this difference would not be significant to examine the usefulness of the proposed framework and the hierarchical model because their use is not much dependent on the specific user interface types or the functions of mobile phones.

Only LUI aspects of mobile phones in this case study were tested because it was thought that it was sufficient only to test LUI for examining the usefulness of the hierarchical model. For this, a provisional checklist to evaluate LUI by referring to the hierarchical model was developed. In total, 11 evaluation criteria were used for evaluating the usability of LUI. Through this case study, it could be confirmed that the checklist reflecting the hierarchical model was useful to evaluate the usability of LUI and to collect usability-related data.

After finishing all the experimental tasks, the participants were interviewed using a questionnaire survey. The questionnaire was designed to examine how the participants perceived the usefulness of the hierarchical model, to obtain their opinions to improve the framework and the model, and to collect the weighting value of tasks, the usability indicator, and the usability property. Most of the participants agreed that the greatest advantage of the hierarchical model is to support usability evaluator's holistic and analytic reasoning to deal with usability-related issues. From the questionnaire survey and informal interview with the participants, the importance of 'pleasure' as a usability indicator and the need of integrating 'learnability' and 'customization' into 'efficiency' was realized. More literature reviews and discussions with usability engineers in mobile phone companies supported them (Dillon, 2001; Hartson, 2003; Jordan, 2000). This point was reflected in the extended model, to be explained later.

The weighting value of tasks, usability indicators, and usability property was used for another study, which is about the quantification of the usability. As this is not the present concern, the weighting value of usability indicators is described briefly. To examine the relative preference of usability indicators, a pair-wise comparison method, which is much used in AHP (Analytic Hierarchy Process) was applied (Saaty, 1994). The weighting values were calculated by using 'geometric mean' of each pair-wise comparison result, and the consistency ratio indicated the obtained values are reasonably reliable. They were: 0.213 for effectiveness, 0.240 for efficiency, 0.190 for learnability, 0.247 for satisfaction, and 0.109 for customization.

Extended Model

Based on the results of the first case study, an extended model was developed by revising the hierarchical model and considering three more issues. The first consideration is the variety of user groups. It is obvious that usability is highly dependent on user-related factors. Moreover, different user groups have their core tasks and specialized tasks. Hence, usability needs to be evaluated differently, taking into consideration the characteristics of a user group. There are several criteria categorizing user groups, such as age, social status, ethnicity, and occupation; however, there is no absolute answer to what is the best criterion. It is thus meaningful to develop different hierarchical models reflecting the characteristics of each user group.

The second consideration is "pleasure" as an indicator. As an information appliance and multimedia device, mobile phones have increasing features that need to be explained from the "pleasure" point of view. The "pleasure" in mobile phones is composed of three benefits: practical, emotional, and hedonic benefits (Jordan, 2000). Practical benefits refer to functional benefits resulted from performing tasks. Emotional benefits occur when a product affects the mood of the users. Hedonic benefits define sensory and aesthetic pleasure associated with the product.

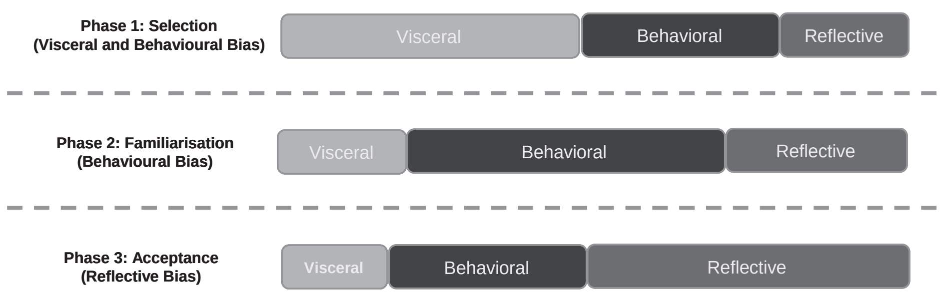
The third consideration is the time varying change of usability, which is emphasized by the dynamic view. In this study, the phases of mobile phone use was divided into three states: selection, familiarization, and acceptance. To reflect such a dynamic view, the three dimensions of design, proposed by Norman (2004), were also adopted. They are visceral design, behavioural design, and reflective design. Visceral design refers primarily to the initial impact and its appearance. Behavioural design is about look and feel from the total experience of using a product. Lastly, reflective design is about thoughts afterwards, how it makes one feel, and the message it tells others about the owner's taste.

Figure 11 illustrates how the three design dimensions are related to the time phases of mobile phone use. At the phase of selection, visceral design aspects have greater impact on usability than behavioural or reflective designs. In contrast, at the phase of acceptance, reflective designs are regarded as the most influential factor. To explain the relationship in more detail, at the phase of selection, users purchasing a new mobile phone and service plan have a visceral/behavioural bias. Examples of the visceral bias are style, finishing, colour, and form. The behavioural bias means that users assess and compare the specification and the functions of the phone and compare them with phones in a similar price range. The phase of familiarization is the critical stage of learning and familiarization with the functions and features of the mobile phone. In this phase, users typically spend an initial period for familiarizing themselves with the functions, features, and interface. According to the informal interview with participants in the first case study, users typically enter the phase of acceptance in a few weeks. If they satisfactorily accept the mobile phone after familiarizing themselves with it, their brand loyalty building on usability is heightened.

The 2nd Case Study

After finishing the first case study, a provisional full set of checklists addressing all five design

Figure 11. Time phases of mobile phone use



areas was developed, referring to the checklist of LUI that was used in the first case study. The second case study was conducted to examine the practicality of the hierarchical model by using the full set of checklists.

Five HCI experts participated in this case study. One expert conducted three tasks (making a call, sending a text, and adding a name and number to the address book) and two experts observed and recorded his actions using a video camera. The other two experts evaluated usability by applying the checklist. After finishing the tasks, five experts discussed the usefulness of the checklist and the hierarchical model, and suggested points that need to be improved. The experience with the two case studies indicated that the approach to identifying and organizing usability factors is viable.

A Set of Checklists

Based on the second case study, a revised version of the checklist reflecting the hierarchical model, which can easily be used for evaluating the usability of mobile phones, was developed. Figure 12 shows a part of the checklist which is concerned with LUI and PUI. This checklist associates usability criteria with their relevant usability properties and metrics, systematically. A more thorough empirical evaluation of the benefits from using this set of checklists remains as a further study.

CONCLUSION

There are a lot of factors affecting the usability of mobile phones. Usability needs to be designed and evaluated, taking account of all the factors in a unified way. A conceptual framework for identifying and classifying usability factors of mobile phones was proposed. The core of the framework is multiple views that address different aspects of interaction between users and mobile phones. They include user view, product view, interaction view, dynamic view, and execution view. The framework intends to be a conceptual

Figure 12a. Checklist based on the hierarchical mode: A part of the checklist to evaluate LUI

Criteria Level	Property Level	Metric						Score
Logical User Interface								
1 Achievement	Operation Sequence was successful (i.e Call was made)	Yes/No	Yes 2				No -1	
2 Flow	Engagement: Time to make call	Time (Secs)	1 ~ 10 4	11 ~20 3	21 ~ 30 2	31 ~ 40 1	> 40 0	
3 Error rate	Input method/ Dedicated key Mapping / Operation Sequence	n (Actions)	0 0	1 ~3 -1	4 ~ 6 -3	7 ~9 -5	>9 -7	
4 Consistency	Actions was repeated successfully	Yes/No	Yes 2				No -1	
5 Navigation	Menu Sturcture		Excellent 4	V.Good 3	Good 2	Average 1	Poor -1	
6 Affordance	Mapping of Soft key with interface		Excellent 4	V.Good 3	Good 2	Average 1	Poor -1	
7 Multi-tasking (time) **Multi-tasking (Bench marked)**	User is asked to perform while performing an additional task (e.g. Undertaking a puzzle)	Time (Secs)	1 ~ 10 4	11 ~20 3	21 ~ 30 2	31 ~ 40 1	> 40 0	
8 Minimalisation (no. of actions) description	*Script:*	n (Actions)	1 ~ 3 0	4 ~ 6 -1	7 ~ 9 -3	10 ~12 -5	> 12 -9	
			Note: the inputting of a number does not count.					
9 Informativess	Feedback		Excellent 4	V.Good 3	Good 2	Average 1	Poor -1	
10 Information Grouping			Excellent 4	V.Good 3	Good 2	Average 1	Poor -1	
11 Quality of Annotation			Excellent 4	V.Good 3	Good 2	Average 1	Poor -1	

Figure 12b. Checklist based on the hierarchical mode: A part of the checklist to evaluate PUI

Criteria Level	Property Level	Metric					Score
Physical User Interface							
1 Power On-Off	Comfort,	Excellent 4	V.Good 3	Good 2	Average 1	Poor -1	
	Spacing	Excellent 4	V.Good 3	Good 2	Average 1	Poor -1	
	Labeling	Excellent 4	V.Good 3	Good 2	Average 1	Poor -1	
	Size	Excellent 4	V.Good 3	Good 2	Average 1	Poor -1	
	Shape	Excellent 4	V.Good 3	Good 2	Average 1	Poor -1	
	Legability	Excellent 4	V.Good 3	Good 2	Average 1	Poor -1	
	Location	Excellent 4	V.Good 3	Good 2	Average 1	Poor -1	
	Haptic Feedback	Excellent 4	V.Good 3	Good 2	Average 1	Poor -1	
	Product Semantics (Approporaite Style for User Group)	Excellent 4	V.Good 3	Good 2	Average 1	Poor -1	

base from which various usability factor models can be developed.

By using three views of the framework, a hierarchical model of usability factors, which has goal-means relationships between adjacent levels, was developed. This model also classifies factors by design areas, such as LUI and PUI. This hierarchical model can effectively be used for several purposes, such as interpreting usability evaluation items and analysing the causes of usability problems. Additionally, an extended model was developed, paying particular attention to the dynamic view of usability and different user groups.

In order to verify the proposed framework and hierarchical model and to obtain information for improving them, two case studies were conducted. These studies supported the usefulness of the proposed framework and model. To support the use of the hierarchical model, a set of checklists for evaluating usability factors by design areas under task scenarios, was developed. It is believed that the framework and model can be useful tools, particularly for expert evaluation and benchmarking.

Although organizing various usability impact factors in terms of goal-means relationships was shown to be useful, the goal-means relationships themselves were not empirically validated. To enhance the practicality of the proposed model, they should be thoroughly examined by conducting a questionnaire survey with large number of users and by applying advanced statistical methods. Lastly, a detailed methodological framework, which can help usability engineers derive a model from the conceptual framework, remains as a matter to be developed.

ACKNOWLEDGMENT

This research was financially supported by LG Electronics Inc.

REFERENCES

Albin, S.T. (2003). *The art of software architecture*. Indianapolis: Wiley.

Bawa, J. (2006). *The particular usability challenge of multifunction devices*. Retrieved January 10, 2007, from http://uk.builder.com

Bevan, N. (1999). Quality in use: Meeting user needs for quality. *Journal of Systems and Software*, *49*(1), 89-96.

Blom, J., Chipchase, J., & Lehikoinen, J. (2005). Contextual and cultural challenges for user mobility research. *Communications of the ACM*, *48*(7), 37-41.

Bosch, J., & Folmer, E. (2004). Architecting for usability: A survey. *Journal of Systems and Software*, *70*(1-2), 61-78.

Dillon, A. (2001). Beyond usability: process, outcome and affect in human-computer interactions. *Canadian Journal of Information Science*, *26*(4), 57-69.

Dillon, A., & Watson, C. (1996). User analysis in HCI: the historical lesson from individual differences research. *International Journal of Human-Computer Studies*, *45*(6), 619-637.

Donyaee, M., Seffah, A., Kline, R., & Striva, J. (2002). An integrated framework for usability measurement. In *Proceedings of the 12th International Conference on Software Quality* (pp. 1-10). Ottwa: ASQ Press.

Fabre, J. B. (2000). *Designing time at the user interface: a study of temporal aspects of usability*. Unpublished doctoral dissertation, Swinburne University of Technology.

Folmer, E., Van Gurp, J., & Bosch, J. (2003). A framework for capturing the relationship between usability and software architecture. *Software Process Improvement and Practice*, *8*(2), 67-87.

Frøkær, E., Hertzum, M., & Hornbæk, K. (2000). Measuring usability: are effectiveness, efficiency, and satisfaction really correlated? In *Proceedings of ACM Conference on Human Factors in Computer Systems (ACM CHI 2000)* (pp. 345-352). New York: ACM Press.

Han, S. H., Yun, M. H., Kwahk, J., & Hong, S. W. (2001). Usability of consumer electronic products. *International Journal of Industrial Ergonomics*, *28*(3-4), 143-151.

Ham, D-H., Heo, J., Fossick, P., Wong, W., Park, S., Song, C., & Bradley, M. (2006). Conceptual framework and models for identifying and organizing usability impact factors of mobile phones. In *Proceedings of 2006 Annual Conference of the Australian Computer-Human Interaction Special Interest Group (OZCHI 2006)* New York: ACM Press.

Hartson, H. R. (2003). Cognitive, physical, sensory, and functional affordances in interaction design. *Behaviour & Information Technology*, *22*(5), 315-338.

Hornbæk, K. (2006). Current practice in measuring usability: Challenges to usability study and research. *International Journal of Human-Computer Studies*, *64*(2), 79-102.

ISO/IEC 15939. (2001). Software engineering—Software measurement process framework.

ISO/IEC 9126-1. (1998). *Software engineering—Software product quality-Part 1: Quality model*.

ISO/IEC 9126-4 (2001). *Software engineering—Software product quality-Part 4: Quality in use metrics.*

ISO/IEC 9241. (1998). *Ergonomic requirements for office work with visual display terminal—Part 11: Guidance on usability.*

Jordan, P. W. (2000). *Designing pleasurable products: an introduction to the new human factors.* London: Taylor & Francis.

Keinonen, T. (1998). *One-dimensional usability: Influence of usability on consumers' product preference*. University of Art and Design in Finland.

Ketola, P., & Röykkee, M. (2001). *The three facets of usability in mobile handsets.* Paper presented at the workshop of ACM Conference on Human Factors in Computing Systems (ACM CHI 2001), Seattle, WA.

Kiljander, H. (2004). *Evolution and usability of mobile phone interaction styles.* Unpublished doctoral dissertation, Helsinki University of Technology, Finland.

Klocktar, T., Carr, D. A., Hedman, A., Johansson, T., & Bengtsson, F. (2003). Usability of mobile phones. In *Proceedings of the 19th International Symposium on Human Factors in Telecommunication*, Berlin: Information Gatekeepers, Inc.

Lauesen, S. (2005). *User interface design: a software engineering perspective*. Essex: Addison Wesley.

Lee, Y. S., Hong, S. W., Smith-Jackson, T. L., Nussebaum, M. A., & Tomioka, K. (2006). Systematic evaluation methodology for cell phone user interfaces. *Interacting with Computers*, *18*(2), 304-325.

Lewis, J. R. (2002). Psychometric evaluation of the PSSUQ using data from five years of usability studies. *International Journal of Human-Computer Interaction*, *14*(3-4), 463-488.

LG Electronics. (2005). *B2050 Handset user guide*. Retrieved December 30, 2005, from http://www.lge.co.uk

Lindholm, C., Keinonen, T., & Kiljander, H. (2003). *Mobile usability: How Nokia changed the face of the mobile phone*. New York: McGraw-Hill.

Ling, C., Hwang, W., & Salvendy, G. (2006). Diversified users' satisfaction with advanced mobile phone features. *Universal Access in the Information Society*, *5*(2), 239-249.

Lynch, K., & Gillmore, S. (2006). *Usability-the key to a product's success.* White Paper of Liminia Application Office. Retrieved January 10, 2007, from http://www.liminia-ao.com

McNamara, N., & Kirakowski, J. (2006). Functionality, usability, and user experience: Three areas of concern. *Interaction*, *November-December*, 26-28.

Nielsen, J. (1993). *Usability engineering.* Oxford, UK: Academic Press.

Norman, D. A. (2004). *Emotional design: Why we love (or hate) everyday things.* Cambridge, MA: Basic Books.

Perlman, G. (1994). Practical usability evaluation. In *Proceedings of ACM Conference on Human Factors in Computer Systems (ACM CHI 1994)* (pp. 407-408). Boston: ACM Press.

Perlman, G. (1998). *Web-based user interface evaluation with questionnaires.* Retrieved December 30, 2006, from http://www.acm.org/~perlman/question.html

Rasmussen, J. (1985). The role of hierarchical knowledge representation in decision making and system management. *IEEE Transactions on Systems, Man, and Cybernetics*, *15*(2), 234-243.

Roth, E. M., Patterson, E. S., & Mumaw, R. J. (2001). Cognitive engineering: Issues in user-centered system design. In J. J. Marciniak (Ed.), *Encyclopedia of Software Engineering* (2nd Ed.). New York: John Wiley & Sons.

Saaty, T. L. (1994). How to make a decision: The analytic hierarchy process. *Interfaces*, *24*(6), 19-43.

Schoeffel, R. (2003). The concept of product usability: a standard to help manufacturers to help consumers. *ISO Bulletin*, *March 5-7.*

Treu, S. (1994). *User interface evaluation: a structured approach*. New York: Plenum Press.

Vicente, K. J. (1999). *Cognitive work analysis.* New Jersey: Lawrence Erlbaum Associates.

Weiss, S. (2002). *Handheld usability.* Chichester, UK: John Wiley & Sons.

Zhang, Z. (2003). *Overview of usability evaluation methods.* Retrieved December 30, 2005, from http://www.usabilityhome.com

KEY TERMS

Abstraction Hierarchy: A multi-level knowledge representation framework for describing the functional structure of a particular work domain or system

Graphical User Interface: Interface related to graphical or visual items presenting information needed for users to conduct tasks

Logical User Interface: Interface concerned with information contents and structure for carrying out tasks

Physical User Interface: Tangible, physical interface supporting physical operation needed for executing tasks

Usability Criteria: Sub-indicator that can be directly measured by at least one specific usability property

Usability Factors: A thing that influences users' perception of usability

Usability Indicator: Hypothetical, abstract conceptual constructs affecting usability that cannot be directly measured

Usability Property: Usability metrics related to the actual user interface of a particular mobile phone

Chapter LIII
Will Laboratory Test Results be Valid in Mobile Contexts?

Anne Kaikkonen
Nokia, Finland

Aki Kekäläinen
TeliaSonera, Finland

Mikael Cankar
TeliaSonera, Finland

Titti Kallio
TeliaSonera, Finland

Anu Kankainen
Idean Research, Finland

ABSTRACT

The phenomena a usability test in the field reveals are different from those uncovered in a classical usability test conducted in a laboratory setting. Comparison studies show that these findings are more related to the user experience and user behaviour than usability and user interaction with the device. Testing in the field is a necessary part of the product development cycle, but the question is what and how to test. Duplicating a laboratory usability test method in the field may not make sense in many cases because the required extra effort does not result in comparable added value, as far as understanding user interaction. Studying user behaviour, on the other hand, requires a less controlled test setting.

INTRODUCTION

The mobile context challenges the user of a mobile system in many ways. The user's attention is divided between interaction with a mobile application and interaction with the environment and other people. The complexity of a real usage environment is a concern for usability practitioners. The question is: Can usability tests conducted in laboratory settings provide results that are valid in real-life mobile contexts?

In this chapter the benefits and drawbacks of mobile application usability testing in laboratory settings and in the field will be discussed. First, the latest views on the nature of the mobile context and how it challenges the mobile user will be presented. Then, some recent discussions concerning usability testing methods in general and issues regarding testing within industry vs. testing with academic goals will be described. After that, studies comparing usability testing in a laboratory with testing in the field, including this study, will be looked at. Some recommendations regarding when to test mobile applications in the laboratory and when in the field, also will be provided.

THE MOBILE CONTEXT

Usability practitioners talk of testing the usability of mobile applications in the field because laboratory settings differ from real usage environments. The mobile context is often considered to be too complex for laboratory simulation. To understand the background, that is, the different aspects of mobility, first the complexity of mobility as a concept needs to be discussed. In the following section, what kinds of challenges mobile users might face when using mobile devices and services on the move, will also be talked about.

Mobility is More than Just being on the Move

The simplest way to think about mobility would be to state that a mobile person is on the move. People travel from place A to place B, visit other places, and wander inside the places (Kristoffersen & Ljungberg, 1999). In reality, we need to remember that people also stop moving and "claim space" for their actions in mobile contexts. For example, people in a bus might pick up a newspaper for some privacy from the surrounding people, or a group of friends who happen to meet each other at a metro station gather in a circle to converse in private, as shown in an ethnographical study by Tamminen et al. (2004). In a sense, people can block out at least parts of their surroundings and concentrate on the task at hand.

Mobile device users may use their devices to build a private environment: When someone needs some privacy in the middle of a busy place, they can take their personal space with the device. A good example could be using a laptop to set up a temporary office, like a "nomadic tent," in a crowded cafe or an airport. The same thing can happen when using a mobile phone. It is not uncommon to see people in public transportation reading, sending messages, or engaging in other activities using their mobile phone. It is a way to gain some privacy.

Mobility on a Larger Scale

Being on the move and stopping to interact with a mobile device does not, however, convey the whole picture of mobility. Kakhira and Sørensen (2002) argue that mobility is not just being on the move but, far more importantly, related to the interaction between mobile people—the way in which people interact with each other in their social lives. Therefore, they suggest expanding the concept of mobility concept by three interrelated dimensions of human interaction: spatial, temporal, and contextual mobility.

Spatial mobility means that not only people, but also objects (such as a mobile phone), symbols (such as news through TV satellites), and spaces (such as virtual communities) move (Kakhira & Sørensen, 2002). Changes in physical contexts are not the only challenge for mobile users, but moving symbols and virtual spaces also require attention and special understanding. This is quite often apparent in usability tests when the user needs to understand, for example, the billing model behind network services or the location of files in virtual spaces.

Temporal mobility is related to how mobile users perceive and use time. Mobile technologies may allow people to speed up time or save it according to their needs. The temporality of human interaction, however, can no longer be explained from a linear "clock-time" perspective alone; it now consists of multiple temporal modes based on each actor's perspective and interpretation of time itself. The increasing temporal mobilization of human interaction is creating new opportunities and constraints for the ecology of social life. (Kakhira & Sørensen, 2002).

Moreover, changing contexts (culture, language, non-verbal communication, environment, other devices, etc.) challenge the mobile user. Contexts in which people act continuously redefine both how they interact with others and with mobile devices. Mobile technologies allow people to interact with each other free from many contextual constrains (Kakhira & Sørensen, 2002). Spatial mobility, on the other hand, requires adapting to constantly changing contexts. For example, a location-based application can engage the mobile user much more in the current context than using e-mail, for example.

The Challenges of Mobility

As seen from the previous section, mobility is not a simple concept. Its complexity has an impact on how mobile users interact with mobile applications and how they experience these applications.

Active Participation

As Kakhira and Sørensen (2002) argue, mobility is frequently psychosocial in nature. Often, mobility also requires active participation (Oulasvirta et al., 2005), which might interfere with mobile applications interaction. An ethnomethodologically oriented mobility study in urban areas conducted by Tamminen et al. (2004) revealed four characteristics of mobile contexts, illustrating active participation.

The first characteristic was that people often have a plan when moving from one place to another, but the plan functions as a framework and leaves space for situational acts such as dropping by, ad hoc meetings, and other forms of "sidestepping." This requires flexibility from the plan and, to a degree, in navigation. It also means that some mobile interaction might be planned (such as reading e-mails on the train), but some of the interaction can consist of ad hoc activities.

The second characteristic the study revealed was that people on the move often solve their navigational problems by interacting with other people. People not only ask other people to give advice on routes or timetables, but they also inform other people about schedule changes, or negotiate what to do next. This is often done using mobile devices. The assumption is that navigational tasks may sometimes be of the highest priority when on the move and other mobile HCI tasks need to release resources for them.

The third characteristic seen in the study was that time plays a crucial role when moving in urban areas. It is often argued that mobile devices free people from the limitations of time and place. Nonetheless, when people are on the move in urban areas, they do face temporal tensions. Some situations may accelerate, so that hurrying and multitasking is necessary, tasks need to be prioritized and some tasks may have to be given up. Sometimes urban mobility requires slowing down or even stopping. For example, missing a bus means you have to wait for the next bus, or arriving early for a meeting, you have to wait for it to start. These temporal tensions may influence the cognitive resources available for interacting with mobile devices.

Finally, the fourth characteristic was that people have a need to multitask while on the move, but mobility may restrict it. For example, opening a door with a key while trying to talk on a mobile phone is challenging as is trying to listen to metro station announcements while talking on the phone. Sometimes, there might also be a need for multiple mobile HCI tasks such as writing a text message and using the calendar, which requires switching between different applications and orientations.

These multiple mobile HCI tasks are easy to simulate in laboratory settings, but tasks involving more than mobile devices are more difficult to simulate as part of a laboratory test environment.

The Competition for Cognitive Resources

Being aware of the environment and tasks related to navigation engages a big part of people's attention and cognitive resources when on the move. Tasks related to this context, like choosing the right bus or metro, or avoiding being hit by a car while crossing a street, are people's primary tasks in an urban environment. While people use all their senses to monitor what is going on around them, visual resources are particularly important for various tasks (as concluded by Lumsden & Brewster, 2003, among others). Estimating the arrival time of the metro, finding a seat, noticing a friend in the same compartment, and getting off the metro on the right station are just a few of the tasks that require visual cues. Using a mobile device or application competes for cognitive resources with the user's natural active participation with the environment and navigation tasks.

Gonzales et al. (2004) studied how information workers manage multiple activities in a normal work environment and were surprised by the high level of discontinuity in the execution of daily tasks. People spend an average of 3 minutes working on any single task before switching to another. Given their limited cognitive resources, it is interesting that people manage these streams at all.

People's attention is even more fragmented when they move around. According to Oulasvirta et al. (2005), this multitasking in the field leads to a depletion of resources available for task interaction and eventually results in the breakdown of fluent interaction. According to their field study, the test users' continuous attention on the mobile device fragmented and broke down to bursts of just 4 to 8 seconds. The users' attention was diverted from the mobile device to the surrounding environment up to eight times during the time it took for a single mobile Web page to load.

USABILITY TESTING

Since the mobile context challenges the user in so many ways, it is understandable that the ecological validity of usability studies has been a hot topic since the early days of mobile devices. This section will review studies that attempt to resolve the differences between testing mobile applications in field and laboratory settings. In order to discuss the differences, first, what is being talked about needs to be defined.

The Principles of Usability Testing

Usability testing based on the thinking-aloud protocol was originally created and presented by K. A. Ericsson and H. A. Simon in their article Verbal Reports and Data (1980) and a follow-up book, Protocol Analysis (1984). The goal of a test is to study end-user *behaviour* regarding the use of an application or a service, not the user's *opinions*. Questions about opinions can be included as a part of a test session, but basically the usability test method is concerned with observing user behaviour and how the user interface of an application matches the human way of thinking and acting. The usability test protocol is described in more detail in several handbooks, for example in Rubin's (1994) Handbook of Usability Testing.

In a usability test, a test user is advised to think aloud while s/he tries to accomplish a given task. The user is asked, for example, to find a piece of information using a search engine. By observing and listening to the user, the facilitator and other observers find out how the user's thinking proceeds and what s/he expects to find in the user interface. All the silent moments in a session, the wrong paths the user chooses, questions, and so forth, indicate problems in the user interface structure, terminology, or navigation.

Usability testing is a qualitative method as opposed to a quantitative method (questionnaires with statistical analysis, etc.). When the need to collect users' opinions arises, other methods such as those used for market research, must be used. In these studies the number of respondents is typically far higher than in usability tests, that is, up to hundreds or even thousands of people.

Usability testing conducted during an industrial product development process is usually not academic research: The goal of a usability test is to improve the system being developed. Sufficient results are often achieved with 5 to 10 users per test iteration, although all problems may not be detected. The goal of academic HCI research is to better understand users' behaviour and interaction models, as well as improve the methods used in product development. In order for the results to be reliable and to help comparison between studies, the number of test users should be higher. In a paper by Faulkner (2003), a minimum of 95% of usability problems were found with 20 users and variation between groups was fairly small.

Usability Testing in Industry

Usability testing can be adapted for different applications. Resources for application development in an industrial context are usually limited, and usability activities such as user-centred design and usability testing, must be performed cost-effectively. The goal of usability testing in product development is to find severe and disturbing usability problems within the strict limitations of project budgets and deadlines. It is rarely possible, or necessary, to remove every minor glitch from a user interface before a product launch.

Since time and resources are critical, companies look into the most efficient ways to find usability problems in products. Sometimes this means taking shortcuts that should not be taken. Indeed, Ramey and Boren (2000) have investigated the practice of testing and found that often the original usability testing procedure is not properly followed. When resources are limited, attention must be paid to expertise in testing. There usually is not much time for trial and error or training.

Wixon (2003) has also raised special issues to take into consideration when testing cost-effectively in the business world. He says that it is not just the usability problems found in the tests that are relevant for product development, but it is necessary to use a testing framework, which defines how the service can be improved in the shortest time with the least effort.

A commonly accepted recommendation in industry environments is that usability tests with only five test users can reveal 85% of user interface problems (basic human cognitive processes vary little) (Nielsen, 2000). Such a requirement can be criticized or evaluated further in an academic sense, but it is a good example of the efficiency demands present in product development projects.

In addition, usability testing in the field is more time consuming than laboratory testing (Kaikkonen et al., 2005; Kjeldskov & Graham, 2003). Without concrete proof to support the theory that testing in a real-life context is significantly better than a laboratory test, companies have good reason to question whether investing in more expensive and more time consuming field tests is worthwhile.

Usability Testing in the Field and in the Laboratory

Since the advent of mobile systems and services, usability practitioners have discussed the ecological validity of laboratory usability studies and how much results could be improved by testing in the field. A controlled environment is far removed from real-life contexts and may lead to biases in test results. Maintaining dedicated usability laboratories is an expense for companies and their very need has also been questioned.

Modern Testing Equipment

It is only recently that truly mobile recording equipment environments have become available to researchers. Unsurprisingly, up to recent times most (71%) mobile device evaluations have taken place in laboratory settings (Kjeldskov & Graham, 2003). Without proper equipment it was impossible to gather field test data the way it is done in a laboratory environment, that is, by following the user's actions step by step and recording them for further analysis. Miniature cameras now allow proper data gathering in a variety of test settings without obstructing the user in his/her performance of the tasks. The dynamics between the moderator and the user can be similar to those found in a laboratory environment.

Once miniature cameras became available, researchers have used them in different ways to study users and services in real-life environments. Roto et al. (2004) show how usability experiments can be conducted in a field environment using a mobile miniature camera for recording not only the user's actions on the mobile device, but also the user's surroundings. They recommend that field usability tests be conducted in situations where user interaction with the environment is investigated, in addition to interaction with the system.

Comparative Field and Laboratory Studies

As technology allowed usability testing to move out of the laboratory, researchers started studying what this meant for their studies. The following will present a cross-section of recent research into the differences of testing in the laboratory and in real-life environments. The papers tend not to define what is meant by the terms *usability* and *usability problem*, making it very difficult to understand how the outcomes actually differ from each other and whether the authors are talking about the same issues in their conclusions. Also, the number of test users is often so small that the variation of findings within a group is likely to be as big as the variation between the groups.

Kjeldskov et al. (2004) conducted a comparison study of an expert application for health care professionals. The study was conducted in a laboratory setting that was built to resemble a part of a physical space in a hospital department. The tasks in the study were related to the daily activities of the hospital personnel. The field test was conducted in an actual hospital environment. In the study, with six test users in the field and six test users in the laboratory, Kjeldskov et al. came to the conclusion that testing in the field adds little value compared to the laboratory test. Molich's classification was used in the analysis, but no clear definition of a usability problem was given. Some problems in this study did not come out in the field setting and the field setting involved events that decreased control over the study.

Duh et al. (2006) also conducted a comparison study between field and laboratory settings. Twenty users participated in this test, 10 in both settings. The laboratory part was conducted at the usability laboratory of a university and the field test was conducted on a train in Singapore. The tasks in the study were related to activities people might engage in while using public transportation, such as normal use of a mobile phone. In the laboratory, the researchers gave the test participants task scenarios to help them understand the actual use context. In this study more critical problems were found in the field test setting than in the laboratory. The reasons for this, according to the researchers, were that different disturbances (noise, movements, lack of privacy), among other factors, affected test user performance. They concluded that some problems are only found in a field environment. Comparison of these results to other studies is difficult as no explicit definition of what constitutes a usability problem is provided. The used definition seems to differ from standards or other studies in a sense that it seems to include users' behavioural patterns and user interaction with the environment.

Holtz Betiol and de Abreu Cybis (2005) performed a comparison test of three approaches: laboratory tests with a PC-based emulator, laboratory tests with an actual mobile device, and field tests. They had groups of 12 users in each case and the tasks the users performed were related to the use of a mobile portal, that is, tasks that are relatively common for ordinary users. The field part of the test was not performed on the move, but participants were placed in a noisy environment. The study is noteworthy because it actually defined what is meant by a usability problem: the authors used the ISO 9241-11 definition. The results in this study did not show statistically significant differences between the laboratory and field tests when the mobile device itself was used.

Baillie and Schatz (2005) explored multimodality in two conditions: they used testing as one part of the application development process. They used a fairly small number of four to six users per setting in their experiment since the test was only one of several methods used to evaluate the system. The researchers were surprised by the results. It took less time to complete the tasks in the field than in the laboratory. More problems were found in the laboratory environment than in the field, even though there was no difference between the environments when it came to critical problems. Again, the definitions for usability and usability problems were not provided in the paper, making it difficult to compare the results with other studies.

Our Study

The authors of this chapter work closely with mobile terminals and services. The details of testing methods are critical to the work. Based on a number of usability tests that were ran in the work, both in the laboratory and in various out-of-laboratory environments, it was felt that the experiences differed somewhat from ones detailed in the studies that were being read. To understand and verify whether testing in the two environments produces different results, two parallel usability evaluations of a mobile application were organised. One test took place in a typical usability laboratory and the other in the field, including tasks like walking in a shopping centre and using the subway (Kaikkonen et al., 2005).

The two test rounds in different contexts were designed to be as similar as possible. The thinking-aloud protocol and the same predefined series of test tasks were used in both cases—the goal was to make sure the context was the only changing variable. The users were prompted to explain what they were doing, what they expected to happen when making selections, and whether something unexpected happened after the selection. One can question whether conducting the same test in such different environments is meaningful, but it was wanted that the test situations be made as similar as possible to find out if changing one variable, the environment, would make a difference.

The total number of test participants was 40, 20 in both settings. Using a relatively large number of participants meant that variations within groups should not be bigger than between groups.

Special equipment allowed the moderators to run and record the tests. A test user carried a backpack with miniature cameras for recording both the mobile device interface and the user (facial expressions). The moderator could follow the camera image, live, from a wireless six inch monitor that also had an additional camera to record the user's surroundings or anything the moderator considered relevant. The test situation can be seen in Figure 1.

The problems found in different test settings were listed and analyzed both quantitatively and qualitatively. This study produced several results, some of which were contrary to the expectations.

Figure 1. Field test ongoing

The main finding was that there was no difference in the number of problems occurring in the two test settings. In fact, the same usability problems were uncovered in both test settings.

Some differences could, however, be seen between the two settings when the frequency of each problem was studied. The problems that occurred more often in the field seemed to be related to understanding the logic of the relatively complicated application. On the other hand, there were also complex issues where no difference between the two test settings could be discerned.

Even individual task execution times in the field were no longer than those in the laboratory. However, the total time needed for the testing was longer in the field because of the preparation necessary for using the field-testing equipment. Some of the test tasks were performed at specific locations to make the tasks more sensible (for example, taking a picture of flowers in a shopping centre), which also contributed to the longer time the test took in the field.

Interesting observations were made about user behaviour that was more related to user experience than usability during the tests. The many interruptions in the field test did not seem to affect the user's performance. Other metro passengers, for example, did not seem to bother users, even if they came to talk to the moderator. In an extreme case, four security guards at a shopping centre were staring at a user whose backpack looked very suspicious, but the user did not even notice the guards. This was not the case with the moderator, who felt quite awkward at the time.

The users in the field tests concentrated intensely on the tasks at hand, and a few users were able to perform all the tasks while walking. Given a more complex task, the users sought a spot where they were safe from surrounding disturbances, essentially creating a bubble of privacy around them. Creating a safe haven in a public place is natural for users, but how much of an impact the artificial nature of the test setup had on the users' cognitive load also has to be considered. The users typically did not have access to their own address books and had only limited experience with using the device being tested. These issues may mean that users are not able to multitask as well as they would when using their own, familiar devices.

Slowing down or even stopping to perform complex tasks is very much in line with the findings of Mizobuchi et al. (2005), who, in a controlled environment, observed that the walking speed of test participants was fairly slow when typing on the mobile phone. This behaviour gave insight into the difficulty level of the tasks and is difficult to observe in a laboratory setting.

SUGGESTIONS FOR FIELD TESTING

Most of the comparison studies presented in the previous section, including the one we conducted, indicate that conducting usability studies of mobile user interfaces in the field is not worth it. In some cases, however, it may make sense to conduct field tests depending on what kinds of user interfaces are being tested and what kinds of usability problems are to be expected. For example, if the intention is to test talking on the phone in noisy environments such as the metro (Duh et al., 2006) and it's not possible to realistically simulate the noise in a laboratory, it makes sense to test in the field.

Location-based and context-aware services are another example. Testing whether people can find the right route using a GPS navigation tool in a laboratory would be difficult, as this depends on how the user succeeds in transferring the map representation to the actual environment.

Tactile feedback is another area that is difficult to study in the laboratory. The difference in the user's attention level may also have implications for how they notice progress indicators in the application. In a real environment users may pick up a newspaper or check their own phone for calls while waiting for the device to finish a download task, whereas in the laboratory they just stare at the phone screen for minutes at a time (Kaikkonen et al., 2005).

Sometimes usability testing requires little additional effort as part of a field trial that is already taking place. During such trials, prototypes of a system being designed are given to test users for use in their everyday life for a longer period of time (such as 4 weeks). During this time the users can be interviewed and observed several times in order to study not only usability issues but also behavioural patterns emerging from interaction with the prototype. Log files can be used to collect additional user data. The result is a deeper understanding of why, how, and in what contexts users would use the system being developed (see for example, Mäkelä et al., 2000). When prototypes are tested with groups of people who interact with each other during the trial, social interactions can also be studied. As Kakhira and Sørensen (2002) point out, mobility also involves social interaction and not only being on the move.

Choosing a Location for Out-of-Laboratory Testing

Mobile phones and other mobile devices are used 'anywhere' and defining a good out-of-laboratory location to test a device or an application is not a simple task. The location used in the test should be one where people normally use mobile devices. Specifically, it should be socially acceptable to use such a device at the test location.

Calling on the mobile phone, for example, may irritate bystanders (Love & Perry, 2004), but there are places where even text messaging is inappropriate. Test users are usually acutely aware of social norms related to phone usage in public places in their own environment and breaking the norms might make the test users feel uncomfortable, like Palen et al. reported in their study (2000). Users should not be given tasks that force them to act against the social norms of the test location.

Diverse places such as cafés, cinemas, transportation, and streets, have different social codes, depending on how they are built. Fyfe (1998) writes about the effect of architecture on people's behaviour in public places and differences in different cities; the way the environment is built either encourages or discourages social communication, walking in the streets, and other behaviour. These kinds of architectural effects need to be taken into consideration when planning the test environment.

When testing in an unfamiliar environment, it would be beneficial to ask local people about norms and social codes, or observe how people behave prior to test planning. This also helps evaluate the validity of the test results. In order to understand how ecologically valid the test situation is, user behaviour related to the test device and service needs to be analysed. And even that is not enough. When running a test in a public place, whether user behaviour differs from the social code in that particular environment also needs to be observed. The ways in which people generally create private spaces in public areas should be understood in order to draw the right conclusions from a test user's behaviour during a test session (Kopomaa 2000).

If a test is conducted in the "wrong" place, the results may give more insight on test user interaction with the environment and other people than with the tested device or service. Testing in a socially unacceptable place may also create unnecessary stress for the user and s/he may not be able to concentrate on other issues.

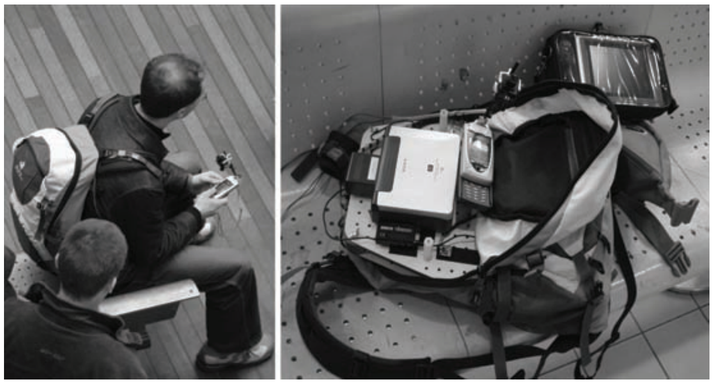
Choosing a test location may also depend on what usage is studied: the initial experience of learning to use a device or later, continued usage. Based on the information with mobile phones and services, for example, people tend to try out new gadgets at home or some other peaceful place, while the eventual usage environment may well be a bus or a crowded restaurant.

The Logistics of Field Testing

The relatively complex equipment necessary for recording user interaction in the field requires more preparations than the familiar equipment used in a laboratory. There are batteries to recharge, the backpack must be adjusted for each user, and explaining how the user should behave during the test typically takes longer. The complexity of the equipment can be seen in Figure 2.

This means that field tests take more time than laboratory tests, as can be seen in table 1 (according to our study). In practice, one can run fewer tests per day in the field than in the laboratory.

Figure 2. Equipment used in field tests



Field tests are vulnerable to unexpected events, such as rain or bus schedules. These risks should be listed before the test is run with actual test users. Since the environment cannot be controlled in the same way as the laboratory, the researchers should also have a backup plan or recruit an extra user, just in case. Running a pre-test or a pilot is critical to the success of a field study. This helps to reduce the risks due to the technology used, but it also helps identify factors that may influence the analysis of the results. If the user moves around during the test, for example, is there a location where the lighting makes it impossible to see the text on a screen, or the surrounding noise blocks out the notifications of the device? If the test focuses on software rather than hardware issues, these kinds of environmental disturbances may make it impossible to get any meaningful results from the test.

There are several test planning issues that must be specified in greater detail for a field test than a laboratory test—particularly if multiple moderators run the test or the tests are outsourced. Examples of these issues are moderator prompting, timing between questions, how to react to external interruptions, and to what extent test user behaviour is controlled. Since the field setting is less predictable, specifying these details takes additional effort.

It is important to be open about the nature of the test when recruiting users. Some users may not be willing to participate when they hear the test will take place in a public location—it happened with a few users. Facing this issue while

Table 1. Differences between locations

	Laboratory	Field
Total test time per user, average	35 min.	45 min.
Instructions and preparations per user, estimated time	10 min.	20 min.
All user interface problems found	Yes	Yes
Users easily understood the application concept	Yes	No
User behaviour can be observed in a natural environment	No	Yes
Environment can be fully controlled	Yes	No
Suitable for usability testing	Yes	Yes
Suitable for testing a concept or service idea	With restrictions	Yes

recruiting is a lot easier than having irate users quitting in the middle of the test.

The effect of the recording equipment on the test user needs to be taken into consideration. Even with miniature cameras, the backpack may be too heavy for some users, possibly limiting the duration of the test sessions. Having the moderator carry as much equipment as possible on behalf of the test user is recommended. If the equipment is conspicuous, the test user may find carrying it embarrassing which may produce a bias in the test results and make it harder to recruit users.

The tasks planned for the field test need to be natural for the test environment. As discussed earlier, the environment and an unfamiliar device increase the users' cognitive load and they may not be able to multitask as well as they would in a normal situation, using a familiar device.

Even after careful preparation, field tests are unique events. Potential interruptions and overall user behaviour need to be taken into consideration when analysing the data from each test.

CONCLUSION

Most comparison studies, including this study, indicate that conducting usability studies of mobile applications or devices in the field in order to find usability problems alone in user interaction with a system, that is, usability problems as defined in the ISO 9241-11 standard, is not worth it. Based on these findings, the recommendation for most testing needs is to use the available resources to perform several quick laboratory tests iteratively during the design process, rather than concentrate efforts on a single field test.

There are also situations where laboratory testing is not enough. In some cases, the limitations of a laboratory setting may be technical. GPS navigation, for example, does not work indoors. Some environmental factors, such as noise, can be difficult to simulate realistically. In other cases the limitation may be a result of how the device is used together with the environment. Again using GPS navigation as an example, a real test task involves mapping information from the device to the surroundings.

Field-testing can also be useful when the purpose is, in addition to testing the usability of a user interface, to gain knowledge about user behaviour in a natural environment, that is, to understand where users might use the service. During the first stages of the product development process, the most important information comes from understanding users and the environments where the service is going to be used. Observation, in-depth interviews, and other methods used in psychology and sociology provide information that better describes the needs of the users, as well as possibilities and restrictions for the service. From a service design point of view, if a service is supposed to be usable while the user is on the move, the designer has to know what "on the move" means for the users of that particular application. In general, it is crucial for product developers to understand the users' usage patterns and multitasking requirements because it helps create better services.

Later on, with prototypes or first versions of the service, evaluation comes in to play, but there are a lot of unanswered questions beyond simple usability problems. Conducting a usability test in the field is one way to find usability problems and get information that can be acquired more easily in the right context. On the other hand, user testing as part of a product development process does differ from user research, even if the methods used can be similar.

Finally, one explanation for similar findings in laboratory and field environments could be that most mobile services require such a high level of concentration, forcing users to create "a bubble" around them and stop other activities. Maybe user interfaces that are easier to use will not only open up new possibilities for using the services in different situations, leading to a better user experience, but also bring opportunities for mobile services and device manufacturers. This would also mean that testing in the field should be re-evaluated if new, easier user interaction models change user behaviour on the move.

REFERENCES

Baillie, L., & Schatz, R. (2005). Exploring multimodality in the laboratory and the field. In proceedings *In Proceedings of ICMI 2005.*

Boren, T., & Ramey, J. (2000). Thinking aloud: Reconciling theory and practice. *IEEE Transactions on Professional Communication*, *43*(3), 261-278.

Duh, H. B.-L., Tan G., & Chen, V. (2006). Usability evaluation for mobile device: A comparison of laboratory and field tests. In *Proceedings of MobileHCI 2006*

Ericsson, K. A., & Simon, H. A. (1980). Verbal reports as data. *Psychological Review*, *87,* 215-251.

Ericsson, K. A., & Simon, H. A. (1984). *Protocol analysis: Verbal reports as data*. Cambridge, MA: MIT Press.

Faulkner, L. (2003). Beyond the five-user assumption: Benefits of increased sample sizes in usability testing. *Behavior Research Methods, Instruments and Computers*, *35*(3), 379- 383.

Fyfe, N. R. (1998). *Images of the street* (Introduction, pp. 1-10). NY: Routledge.

González, V. M., & Mark, G. (2004). Constant, constant, multi-tasking craziness: managing multiple working spheres. In *Proceedings CHI'04* (pp. 113-120). ACM Press.

Holtz Betiol, A., & de Abreu Cybis, W. (2005). Usability testing of mobile devices: A comparison of three approaches. In *Proceedings of Interact 2005.*

ISO 9241-11 (1998). *Ergonomic requirements for office work with visual display terminals (VDT)s – Part 11: Guidance on usability.*

Kaikkonen, A., Kekäläinen, A., Cankar, M., Kallio T., & Kankainen A. (2005). Usability testing of mobile applications: A comparison between laboratory and field testing. *Journal of Usability Studies*, *1*(1), 4-16.

Kakhira, M., & Sørensen, C. (2002, January 7-10). Mobility: An extended perspective. In *Proceedings of the Hawaii International Conference on System Sciences*, Big Island, Hawaii.

Kjeldskov, J., & Graham, C. A. (2003). Review of mobile HCI research methods. In *Proceedings of the 5th International Mobile HCI 2003 Conference*, Udine, Italy. Sringer-Verlag, LNCS.

Kjeldskov J., Skov, M. B., Als, B. S., & Høegh, R. T. (2004). Is it worth the hassle? Exploring the added value of evaluating the usability of context-aware mobile systems in the field. In *Proceedings of the 6th International Mobile HCI 2004 Conference* (pp. 61-73). Glasgow, Scotland. Berlin: Springer-Verlag.

Kopomaa, T. (2000). *City in your pocket. Birth of the mobile information society.* Helsinki: Gaudeamus.

Kristoffersen, S., & Ljungberg, F. (1999). Mobile use of IT. In T. K. Käkölä (Ed.), *Proceedings of the 22nd Information Systems Research Seminar in Scandinavia Conference* (Vol. 2, pp. 271-284).

Love S., & Perry, M. (2004). Dealing with mobile conversations in public places: Some implications for the design of socially intrusive technologies. In *Proceedings of CHI 2004.* Vienna: ACM.

Lumsden, J., & Brewster, S. (2003). Paradigm shift: Alternative interaction techniques for use with mobile & wearable devices. In *Proceedings of CASCON'03.*

Mizobuchi S., Chignell, M., & Newton, D. (2005). Mobile text entry: Relationship between walking speed and text input task difficulty. In *Proceedings of the MobileHCI 2005.* Salzburg, Austria: ACM.

Mäkelä, A., Giller, V., Tscheligi, M., & Sefelin, R. (2000). Joking, storytelling, artsharing, expressing affection: A field trial of how children and their social network communicate with digital images in leisure time. *In Proceedings of CHI 2000* (pp. 548-555).

Nielsen, J. (2000). *Why you only need to test with 5 users*. Retrieved January 15, 2007, from http://www.useit.com/alertbox/20000319.html

Oulasvirta, A., Tamminen, S., Roto, V., & Kourelahti, J. (2005). Interaction in 4-second bursts: The fragmented nature of attentional resources in mobile HCI. In *Proceedings of SIGCHI Conference on Human factors in Computing Sytemes* (pp. 919-928). NewYork: ACM Press.

Palen L., Salzman M., & Youngs, E. (2000). Going wireless: Behavior & practice of new mobile phone users. In *Proceedings of CSCW*. Philadelphia: ACM.

Roto, V., Oulasvirta, A., Haikarainen, T., Lehmuskallio, H., & Nyyssönen, T. (2004, August 13). *Examining mobile phone use in the wild with quasi-experimentation* (HIIT Tech. Rep. 2004-1).

Rubin, J. (1994). *Handbook of usability testing*. New York: John Wiley & Sons.

Tamminen, S., Oulasvirta, A., Toiskallio, K., & Kankainen, A. (2004). Understanding mobile contexts. *Personal and Ubiqitous Computing, 8*, 135-143.

Wixon, D. (2003). Evaluating usability methods. *Interactions*, *July-August*, 29-34.

KEY TERMS

Field Test: Usability test in a real-life context

Laboratory Test: Usability test in a controlled environment

Usability: The extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency, and satisfaction in a specified context of use (ISO 9241-11).

Usability Laboratory: A controlled environment where evaluators set up usability tests and other experiments. In a usability laboratory all factors of the tested system can be controlled and high-quality data collection (video, etc.) is possible.

Usability Problem: Problems that influence the effective, efficient, and satisfactory use of the system in a specified context of use (ISO 9241-11)

Usability Test: User test that tests the effectiveness, efficiency, and satisfaction of the system in a specified context of use (ISO 9241-11) using known usability test protocols

User Experience: User's holistic experience with the product-user experience is an intra-user event which is the consequence of how well the product matches with user expectations, how well it supports the activities in different physical and social contexts. The entire user experience may not be possible to detect by using usability testing alone.

Chapter LIV
Mobile Evaluations in a Lab Environment

Murray Crease
National Research Council of Canada, Canada

Robert Longworth
University of New Brunswick, Canada

ABSTRACT

The evaluation of mobile applications is increasingly taking into account the users of such applications' mobility (e.g., Mizobuchi, Chignell, & Newton, 2005; Mustonen , Olkkonen, & Hakkinen, 2004). While clearly an important factor, mobility on its own often does not require the user's visual focus to any great extent. Real-life users, however, are required to be aware of potential hazards while moving through their environment. This chapter outlines a simple classification for describing these distractions and two evaluations into the effect visual distractions have on the users of a mobile application. In both cases, the participants were required to monitor both their environment and the display of their mobile device. The results of both evaluations indicated that monitoring the environment has an effect on both task performance and the subjective workload experienced by the participants, indicating that such distractions should be considered when designing future evaluations.

INTRODUCTION

Mobile computing devices are becoming increasingly popular but the evaluation of such devices has not developed at the same rate. Many early evaluations were undertaken on desktop emulators, often because the real devices were not capable of supporting the applications being evaluated. As the availability and power of mobile devices grew, so too did their use in evaluations, but these evaluations were often still run in a static laboratory environment carefully devoid of distractions. Gradually, however, mobility has become an increasingly common component in the evaluation of mobile devices. Clearly, the fact that a user is likely to be mobile is the single greatest difference in context between users of mobile and desktop devices. This mobility, however, leads to dynamic changes in users' context which may mean the users are not capable of solely focussing on the task at hand. Users may, for example, have to use their visual focus to navigate or they may be listening to a conversation while receiving audio feedback from their device.

This chapter describes two evaluations investigating how visual distractions affect the task performance of a mobile user. In the first evaluation participants were required to navigate a virtual 'maze' using different forms of navigational cues. While navigating through the maze, the participants were required to monitor projections on either side of them. In the second evaluation the participants were required to monitor the display of their wearable computer while moving through a lab and monitoring projections in front of them. The design of both evaluations highlighted the benefits in making the experimenter mobile while running the evaluation. The results showed that forcing the participants to monitor their environment had an impact on the results and should, therefore, be considered in the design of all mobile evaluations.

The remainder of this chapter is structured as follows. The second section gives an overview of the state of the art in mobile evaluations while the third section introduces a classification that can be used to describe the different forms of distraction that can affect mobile users. The fourth and fifth sections describe two experiments that evaluated the effect of visual distractions on users of mobile applications while the sixth section discusses the results of these evaluations in terms of their experimental design in general, and the distractions used in particular.

MOBILE EVALUATION

There are many examples of mobility having an effect on a user's task performance. Brewster, for example, showed that the amount of data entered using button presses was significantly reduced when comparing a seated, indoor user with a mobile, outdoor user (Brewster, 2002). It was also found that the subjective workload experienced by the participants was significantly increased. Brewster suggests that this is not surprising and goes on to say that further research is required to develop appropriate evaluation techniques for the evaluation of mobile devices in realistic situations. This section presents previous research that has incorporated mobility into the evaluation of mobile applications; and in particular, where it has been used in lab evaluations.

Mustonen et al. (2004) investigated the effect of walking on the legibility of mobile phone text. Four walking conditions—natural speed in a corridor, natural speed on a treadmill, fixed speed of 1.5 km/h on a treadmill, and fixed speed of 3 km/h on a treadmill—were compared to determine if the effect of mobility varied with speed. It was found that although mobility had an effect on legibility when reading normal text, there was no significant effect when parsing pseudo-text with a view to finding a text pattern. The overall workload of both tasks, as measured by NASA TLX ratings (Hart & Staveland, 1988), was significantly effected by mobility.

Mizobuchi et al. (2005) investigated the effect of walking on text input. Participants were required to enter English language sayings using one of four sizes of soft keyboard when either stationary or walking along a corridor. The size of the keyboard had a significant effect on the text input speed but walking had no significant effect on the speed. Furthermore, walking only had a significant effect on the number of errors when the participants were using the smallest keyboard. It is suggested that these results indicate that text input and walking can, in general, be viewed as separate tasks that have no effect on each other apart from a fixed cost to each task due to the presence of the other. This was indicated by a reduction in walking speed when inputting text and a reduction in input speed when walking, although these effects were not significant. It could also be argued that the inputting and walking tasks were such that, other than when using the smallest keyboard, the participants had sufficient cognitive and visual capacity to successfully manage both but if the load was increased (e.g., when the keyboard was very small or if more complex navigation was required) then a noticeable effect may become apparent.

Crossan, Murray-Smith, Brewster, Kelly, and Musizza (2005) described a quantitative approach to measuring the effect of walking on usability. Using an accelerometer attached to the serial port of a PDA, it was possible to determine the users' gait while walking and consequently that the rhythm of walking affected the users' ability to select on-screen targets. Participants were required to tap targets that appeared on the screen at random intervals. This was done both when walking and when seated with, not surprisingly, far greater accuracy achieved when seated. In the walking condition it was found that there was a correlation between the phase of the participants' gait and both the accuracy and number of taps.

The examples of mobility in the evaluations presented thus far have been in controlled experimental scenarios where the participants are required to be mobile but are not required to monitor their surroundings as would be the case in a real-world scenario. Kjeldskov and Stage (2004) compared six techniques which could be used to increase the realism of such evaluations: sitting at a table; walking on a treadmill at a constant speed; walking on a treadmill at a variable speed; walking at a constant speed on a course that is constantly changing; walking at a variable speed on a course that is constantly changing; and in a field evaluation (walking in a pedestrian street). The five lab-based techniques related to the five possible combinations of motion (none, constant, and variable) and attention required navigating (yes or no). The different techniques were compared in terms of the number of usability problems found, the task performance, and the subjective workload experienced by the participants. Interestingly, the participants were best able to find usability problems when sitting at a table. It is suggested that this is because the participants are able to devote the most attention to the means by which problems were recorded—thinking aloud. Mobility had no significant impact on the task performance of the participants. It did, however, have an impact on the workload experienced by the participants. Simply being mobile, however, was not sufficient since as walking on a treadmill at a constant speed did not significantly increase the workload experienced. For this to happen, an additional cognitive load was required whether it be variable speed, variable course, a combination of the two, or being in a real-world situation such as a street. Although the overall workload was significantly increased when a variable course was employed, there was no significant increase in mental demand which would be expected due to the increased demand in following a variable course. It was hypothesised that this was due to the way the variable course was managed, with the participants required to follow an experimenter who followed a variable path. This enabled the participants to merely follow the experimenter with no real effort required to manage the navigation.

Kjeldskov and Stage also experienced difficulties in collecting data when running the experiments in the field. It was hard to video the participants as they moved through the streets and the realism of the setting was compromised as other pedestrians tended to avoid the participant and the experimenters. Goodman, Brewster, and Gray (2004) report that a further problem with field studies is the difficulty in controlling confounding variables. They suggest that this problem may be minimised by removing data where it varies too greatly from appropriate control levels. This, however, can be both expensive and time consuming so Goodman et al. suggest that such results should be included as they are part of the real-world context in which the evaluated system is expected to operate.

Kjeldskov, Skov, Als, and Høegh (2004) investigated whether the added effort required to undertake a usability evaluation in the field is worth it in terms of the results such an evaluation produces. The investigation compared the effectiveness of two evaluations of a mobile Electronic Patient Record (EPR) system. Two forms of the evaluation were undertaken: one in a lab-based simulation of a hospital ward and the other in a real hospital ward. The comparison of the two forms of evaluation was based on the number of usability problems identified. Surprisingly, significantly more serious and cosmetic problems were discovered in the lab-based evaluation than in the field-based evaluation. Only one problem discovered in the field was not discovered in the lab and this problem was not even directly linked to the usability of the system but rather to the veracity of the data entered into the system and its storage in a database. Furthermore, it was found that running the experiment in the field posed challenges with respect to the collection of data. Participants in the lab, for example, were prepared to use a note taking facility to document problems they found whereas nurses operating in a real life context (not surprisingly) did not. Although the particular context of this study—a hospital ward where patient safety is the most critical factor—may have been a factor, the results do indicate that if the real-world context is taken into account when creating a lab environment, a lab evaluation may be at least as good as a field evaluation.

Duh, Tan, and Chen (2006) also undertook a comparison of lab and field evaluations. Two groups of participants undertook an evaluation of a mobile-phone based application in one of two settings: seated in a lab with the usage scenario textually described or in the field in the actual usage scenario. In both cases, the think aloud technique was used and the participants' interaction with the application was recorded. In contrast to Kjeldskov et al. (2004), significantly more critical errors were found by the participants in the field than by those in the lab. Although no definitive reason is given, there are several possibilities. The lab-based participants were seated during the evaluation so no attempt was made to mimic the real-life context of use. Also, the participants in the field expressed increased nervousness and stress which may have been an experimental artifact caused by the requirement to verbally describe everything they were doing in a public location.

This section has described some of the work that has been undertaken on making evaluations of mobile devices more realistic and, as a consequence, more effective. The majority of this work, however, has concentrated solely on mobility. Mobility on its own, however, does not require much of the participants' attention whereas in a real-world context, users are required to manage the consequences of mobility—a dynamically changing environment—which do require more attention. One way to solve this problem would be to undertake evaluations in the field but this brings its own set of problems such as difficulties in controlling the environment and capturing data. The remainder of this chapter describes two evaluations that attempted to produce a more realistic environment through the use of controlled, visual distractions.

DISTRACTIONS

As previously discussed, mobility alone is not sufficient when evaluating mobile applications. Real-life users of such applications are required to constantly monitor their environment while moving through it to avoid hazards such as other pedestrians or lamp-posts. Clearly, this monitoring will have an impact on both the attention and cognitive load the user is able to devote to the mobile application. Furthermore, in a real-world context, users will be distracted by the sights, sounds, and smells of the environment regardless of whether these distractions pose a potential risk to the user. This section proposes a simple classification of the different forms of distraction that could be used in lab-based evaluation of mobile devices.

- **Passive distractions** distract users but require no active response. A real-life example of such a distraction is a billboard.
- **Active distractions** require the user to respond in some way. The required response will vary according to the distraction. A mobile phone ringing, for example, may require the user to answer it whereas the presence of a lamp-post will require the user to navigate around it.
- **Interfering distractions**—which may be passive or active—interfere with the user's interaction with the mobile device. The sound of a passing car or an ongoing conversation,

for example, may limit a user's ability to correctly perceive audio feedback being presented by the mobile device. A distraction may be interfering in one context (e.g., the example given) but not in another (e.g., if the user was not relying on audio feedback).

Clearly, in this scheme, the classification of some distractions is subject to debate. A mobile phone ringing, for example, is only an active distraction if the user chooses to answer the call. These decisions, however, can be controlled in an experimental setting (e.g., by instructing users to answer all calls). In this way it is hoped that the effect of different types of distraction and different techniques for managing distractions (e.g., answer all calls immediately or ignore most calls) may be investigated. The remainder of this chapter describes two evaluations that investigate the use of active, visual distractions that interfere with the participants' ability to interact with the visual interface of a mobile application.

EVALUATION 1: COMPARISON OF AUDIO AND VISUAL NAVIGATIONAL CUES

A lab evaluation investigated whether simple, non-spatialised sounds could be effective at enabling users to navigate. Such sounds would have the advantage of not requiring headphones which may occlude environmental sounds of interest to the user.

Background

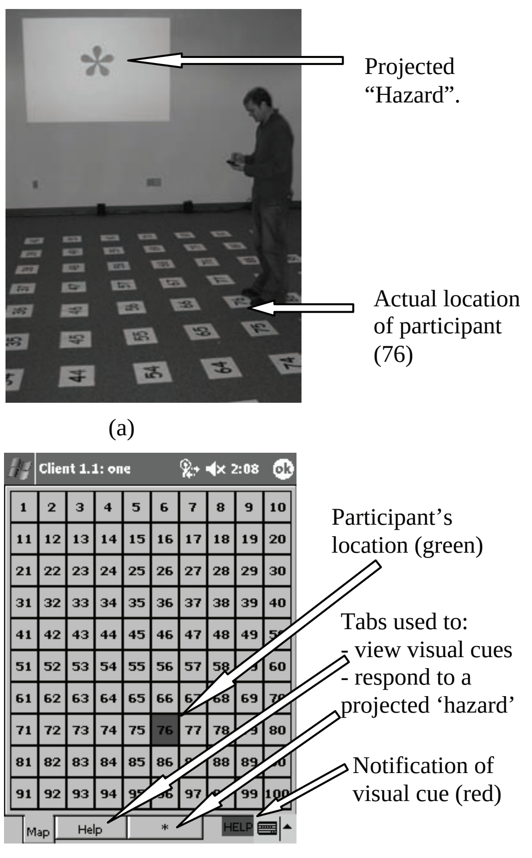
There are several systems which have used audio feedback to provide users with navigational information. MOBIC (Petrie, Johnson, Strothotte, Raab, Fritz, & Michel, 1996) is an example of a system designed to allow blind users to navigate. MOBIC used GPS information mapped to speech that provided users with assistance for macro-navigation—the navigation through the distant environment—which is typically done using visual cues such as church steeples. One of the main findings of the user analysis undertaken was that the users did not want to wear headphones as it was felt that this would block out useful environmental sounds which are especially important to visually-impaired users. More recently, the Personal Guidance System (Loomis, Marston, Golledge, & Klatzky, 2005) was evaluated using different forms of spatial display. As with MOBIC, the system mapped GPS information into direction and distance feedback which was presented to the users sonically.

The use of audio navigation cues is not limited to systems designed for visually impaired users. AudioGPS (Holland, Morse, & Gedenryd, 2002) was designed to allow sighted users to receive navigational information using spatialised non-speech audio presented using headphones. Non-speech sounds were chosen to minimise any interference with the users' conversations. The system presented the users with two pieces of information: distance to their destination (or intermediate waypoint) and its direction. Initial trials indicated that the sounds were effective in allowing users to discern the direction of the destination but that the implementation of the system meant that it was slow to respond to a user's change in direction. The gpsTunes system (Strachan, Eslambolchilar, Murray-Smith, Hughes, & O'Modhrain, 2005) modified currently playing music to provide direction and distance information to the user. As the distance to the target decreased, the volume of the music was increased up to a user defined maximum. The stereo pan of the music indicated the correct direction of the target. When the target was reached, a pulsing sound was played over the music. As with AudioGPS, the sounds were presented using headphones. An initial field evaluation indicated that users were able to navigate successfully using the system but it is unclear whether the annoyance of modifying the music would prove to be annoying to users in the longer term.

Experimental Design

An experiment was designed to investigate two main questions: what affect do visual distractions have on users of mobile devices; and how effective are simple, non-spatialised sounds at providing navigational cues? The experimental task was in the form of a game with the participants required to navigate through a virtual 'maze' as quickly as possible. The participants could not see the maze but only the numbered grid in which the maze was located (Figure 1). As the participants moved through the maze they were provided with a cue indicating in what direction they were to move next (forwards, backwards, left, or right). When they had moved, the participants clicked on their new location in the client interface (Figure 1b) and the task proceeded until the participant reached the end of the maze.

Figure 1. Two views of the same maze: (a) as represented by a grid of cells on the floor through which the participants must navigate; and (b) as it is presented to the participants on the handheld device

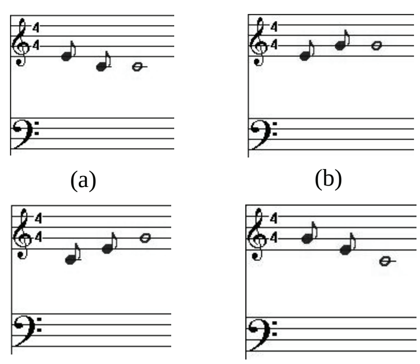


Two forms of navigational cue representing the directions left, right, forward, and backward were used. Visual cues took the form of an arrow pointing in the appropriate direction. These cues were accessed by pressing the "Help" tab at the bottom of the screen (Figure 1b). The arrows were visible for 350ms before the display returned to the view of the map. The arrow could only be viewed once for each step. 350ms was chosen as a suitable length of time to display the visual cue as Öquist and Goldstein (2001) report that the average time required to fixate—or parse—visual information is ~330 ms.

Four earcons (Blattner, Sumikawa, & Greenberg, 1989) which all shared the same basic structure—two notes of duration 80ms followed by a third note of duration 480ms—were used for the audio cues. All the notes were played in a piano timbre. The directions were differentiated using the notes' pitch:

- Left—E3, followed by C3, followed by C3[1] (Figure 2a)
- Right—E3, followed by G3, followed by G3 (Figure 2b)
- Forward—C3, followed by E3, followed by G3 (Figure 2c)
- Back—G3, followed by E3, followed by C3 (Figure 2d)

Figure 2. The four sounds: (a) left; (b) right; (c) forward; and (d) backward



By using pitch as the only parameter by which the sounds can be distinguished, the possibility for encoding more complex navigational cues which could inform the user about a more detailed direction and/or the distance to be travelled is left open. By relying only on relative pitch (i.e., the way the pitch changes within the earcons) as opposed to relying on absolute pitch (i.e., the way the pitch changes between earcons), these sounds follow guidelines on the design of audio feedback (Lumsden, Brewster, Crease, & Gray, 2002). Because the earcons are not spatialised, the user was not required to wear headphones which may block out other sounds (Petrie et al., 1996).

To simulate a realistic mobile environment—where users are required to be aware of their surroundings—the participants were also required to monitor their surroundings and react to distractions accordingly. Projections were displayed at pseudo-random intervals on either side of the participants. Figure 1a, for example, shows a '*' being projected to the right of the participant. Two forms of projection were used: characters and images. In the characters condition, six characters were used to represent non-hazards—'u,' 'v,' 'w,' 'x,' 'y,' 'z;' a seventh—'*'—was used to represent a hazard. In the images condition, seven images replaced the seven characters of the characters condition. The non-hazard images were pictures of empty roads. The hazard image had a single moving car on a road. Participants were required to respond to the projection of a 'hazard' by pressing a tab at the bottom of the interface on the handheld (Figure 1b).

Experimental Procedure

The experiment consisted of three conditions run between three groups. The three conditions were: visual cues only; audio cues only; and both audio and visual cues. The three groups were: character distractions; image distractions; and no distractions. The third group was used as a control group to determine whether the participant monitoring their environment had an effect on the evaluation results. Twenty-four participants were divided equally between the three groups. The order in which the audio and visual conditions were presented within each group was counterbalanced with the audio-visual condition always presented third. This design allowed the effectiveness of the two different forms of cues to be determined. The audio-visual condition allowed us to quantitatively determine whether the participants chose to use the audio or visual cues after experiencing both. This was measured by recording how often the participants clicked on the 'Help' tab to view the visual navigational cues.

Each condition consisted of a short training session where the participants were able to familiarise themselves with the navigational cues for that condition followed by a short training maze consisting of 16 steps. The participants then had to navigate a 40 step maze for the actual condition. Three full length and three training mazes were designed; each with an equal number of forward and backward steps (10 of each for the full length mazes and four of each for the training mazes). This was done to ensure the mazes were all of similar difficulty as pilot participants found it harder to move backwards than forwards. To eliminate the mazes as an experimental variable every participant was required to navigate the same mazes in the same order regardless of condition. Similarly, the same projection sequences were used in the same order for every participant.

Results

The different forms of visual distractions used had no significant effect on the participants' performance in the navigation task. The different forms of distraction did, however, effect the participants workload with a two factor ANOVA showing that distraction type significantly effected time pressure ($F_{2,68}$ =5.72, $p<0.01$), performance level achieved ($F_{2,68}$ =7.65, $p<0.01$), and the overall workload ($F_{2,68}$ =4.14, $p=0.021$) experienced by the participants as shown in Table 1.

Post hoc Tukey HSD tests showed that the use of image distractions significantly increased the workload compared to the no distractions group (time pressure $p<0.01$, performance level $p<0.01$, overall workload $p<0.02$). There were no

Table 1. Average workload results across the different distraction groups

Workload / Distraction	Average Time Pressure	Average Performance Level Achieved	Average Overall Workload
Character	8.96	14.08	44.94
Image	11.9	11.87	51.19
None	7.29	16.67	37.56

significant differences in the subjective workload between the character distractions group and the other groups.

A two factor ANOVA showed that the different distraction types significantly affected the participants' ability to correctly detect projected hazards ($F_{2,68}$=10.68, p<0.01). *Post hoc* Tukey HSD tests showed that the participants were significantly better at detecting hazards in the character group—67% correctly noticed—compared to the image group—40% correctly noticed (p<0.01). This result was confirmed by a significant increase in the number of correct responses in the character group—on average 8.25 per condition—compared to the images group—on average 4.83 per condition (p<0.01).

When analysing the use of the audio cues the audio-visual condition was ignored due to the likely training effect caused by this condition always being third. Two analyses were undertaken, one across the two distraction groups and one across all three groups.

A two factor ANOVA test across the two groups with distractions, showed the average time taken to make a correct step was significantly reduced from 5.5 seconds in the visual condition to 3.9 seconds in the audio condition ($F_{31,1}$=21.27, p<0.01). The overall task time was not significantly reduced, however, as there were more navigational errors made in the audio condition (4.5) than in the visual condition (1.1). This, however, was not a significant difference ($F_{31,1}$=2.99, p=0.09).

When considering the subjective workload, a paired T-test showed that the physical workload experienced by the users in the audio condition was significantly reduced from 3.9 on average in the visual condition to 2.6 (T_{15}=2.4, p<0.04) (Figure 3).

When similar comparisons were performed across all the groups the improvement in the time taken to take a correct step when using the audio feedback was maintained. When considering the subjective workload, however, significant increases in mental demand ($F_{1,47}$=15.56, p<0.01), effort expended ($F_{1,47}$=5.52, p<0.03), frustration ($F_{1,47}$=5.01, p<0.04), and overall workload ($F_{1,47}$=15.56, p<0.01) were found in the audio condition.

Figure 3. Average physical demand experienced by the participants who were monitoring distractions

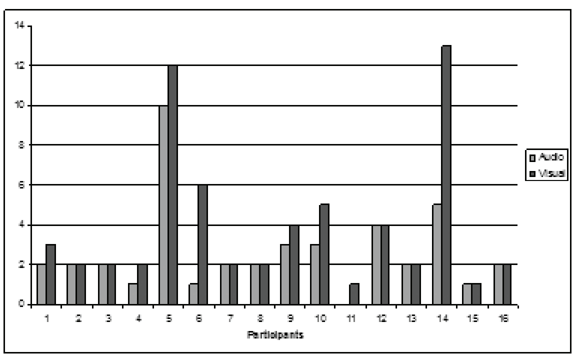


Table 2. Percentage use of visual help in the audio-visual conditions across the three distraction groups

Distraction Type	% Use of Visual Help
Character	20.78%
Image	29.40%
None	40.84%

The audio-visual condition was subjectively analysed to see how often the participants chose to use the visual feedback when they had the option (Table 2).

Discussion

Overall, the audio cues were not wholly successful. Although the average time to make a complete step was significantly improved the overall task time did not reflect this. This was almost certainly due to the increased number of navigational errors made in the audio condition. Although not a significant increase, the cost of an error (each required up to 5 extra steps to be made) meant that any benefit in the time taken to make a step was lost. The improvement in the time to take a step can almost certainly be attributed to the fact that the audio cues were automatically presented to the participants whereas the visual cues had to be actively retrieved.

Anecdotally, however, the audio cues did seem to be preferred by many users with one in particular commenting "I found the visual to be the hardest even though I expected it to be easier." Other common comments among those who preferred the audio cues were that they allowed the participants to monitor their surroundings more easily, with the visual cues requiring too much of their visual attention. Conversely, many participants commented that they found the sounds hard to differentiate and therefore unsuitable for the task at hand. In these instances, the participants were most likely to fail to differentiate the sounds representing forward/right and/or back/left. This was especially true if the participant had moved several steps in the same direction before being required to change to the direction they found hard to differentiate. Observing the evaluations, it became obvious that many participants struggled with the audio cues initially but eventually reached an understanding of how the sounds were structured. Some participants, for example, navigated poorly in the training maze for the audio condition but performed flawlessly in the audio condition itself, indicating that despite having an opportunity to learn the sounds prior to the training maze, perhaps a longer training maze would have been appropriate.

Of more interest is the interaction between the distractions and the audio cues which was perhaps best highlighted by the number of times the visual cues were used in the audio-visual condition, Table 2. The visual help was only used 21% of the time when character distractions were used, 29% of the time when images were used, and 41% of the time when no distractions were used. In the latter case it is likely that, because the users were not required to monitor their surroundings, their visual focus could remain on the PDA and so the effort required to parse the sounds was not as necessary. The difference between the character and image distractions could be explained in terms of the participants' increased difficulty in determining what a hazard in the image condition was. The extra effort required to spot the hazards meaning the participants could spare less effort to parse the sounds.

It was also noticeable that, when comparing the overall workload in the audio condition across all three groups, it was significantly higher in the audio condition than in the visual condition. If only the two groups that had visual distractions were considered, then the only significant difference in workload was an increase in physical demand in the visual condition. This indicates that when the distractions were present, the audio meant the participants only had to turn left and right to monitor their surroundings while without the audio the participants were required to monitor the iPaq also. If no distractions were present, however, this extra physical demand was not present, meaning the extra effort required to parse the audio feedback became a significant factor.

The number of character hazards successfully detected was significantly higher than the number of image hazards. This was somewhat surprising given that in both instances the participants were looking for a single projection out of seven possibilities. In one case it was a '*' and in another it was the picture of a car on a road. The characters projected were chosen to be as similar as possible (u, v, w, x, y, z, and *) as they largely all consisted of diagonal straight lines. It may have been that the '*' stood out due to the density of lines? The images were also all similar, containing a road disappearing into the horizon but perhaps the inclusion of the car was not as noticeable as the '*'.

EVALUATION 2: COMPARISON OF WEARABLE DISPLAYS

While mobile computing is typically associated with hand-held devices such as PDAs and cell phones, wearable computers are becoming increasingly common due to several advantages they possess: speed of access; hands free use; and privacy (Starner, 2003). Typically, wearable displays take the form of head or glasses-mounted displays (HMD or GMD) which utilise small displays located close to the eye enabling a high resolution image to be shown. Displays may be visible in one eye (monocular) or two (binocular). They may be transparent—enabling the user to see the surrounding environment through the display —or opaque. An alternative form of wearable display is a standard flat panel screen that is worn by the user—typically attached to a belt. The advantages of such displays include larger screen size and less occlusion of the environment. The main disadvantages are: the screens can be awkward to wear; reduced privacy; and users are required to actively move their visual focus away from the environment to the screen. This section describes a lab evaluation evaluating the effectiveness of two different forms of wearable display (a fold-down screen and a GMD) under different conditions.

Background

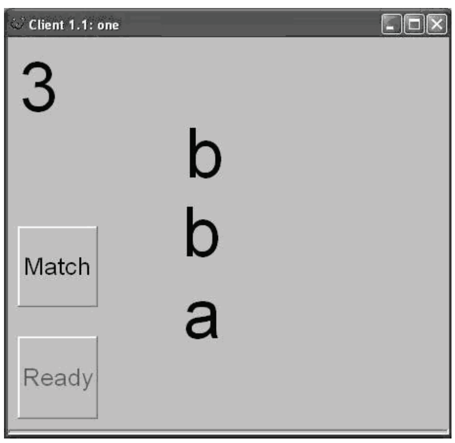
Sheedy and Bergstrom (2002) evaluated the effectiveness of different HMDs compared to different forms of monitor and hard copy for reading tasks. They found that the participants' performance was comparable across all displays. This result differed from previous research where the performance of participants using head-mounted displays was not comparable. It is suggested that these differences may have been caused—in part—by a lack of movement in the evaluation. Revels and Kancler (2000) undertook an evaluation of head-mounted displays that did incorporate mobility. Participants were required to perform three different tasks using a GMD while navigating one of two courses. All the tasks required the participants to press an appropriate button on a wrist mounted keypad based on a visual cue. Three forms of visual cue were used: graphical (a graphical representation of the keypad was shown with the appropriate key highlighted); numerical (the numerical label of the key to be pressed was shown); and textual (a textual description of the location of the key to be pressed was shown). The two courses were: clear (a straight course 100 feet long); and obstructed (participants were required to slalom through a series of different obstacles on a 100 foot long course). The results indicated the participants found the textual task the hardest, followed by the numerical and graphical tasks. The impact of having to avoid obstacles also significantly affected the participants' performance. Most interesting, however, was the fact that in the clear course condition the time taken to respond to a cue was significantly different between all three tasks whereas in the obstacle course condition only the graphical task was significantly different from the other tasks. This indicates the importance of incorporating a realistic environment in mobile evaluations in order to achieve the most accurate results.

Experimental Design

A lab evaluation was undertaken to determine the effectiveness of two common forms of wearable display: a glasses mounted display and a fold down screen. The experimental application ran on a Xybernaut MAV wearable computer running Windows XP. The fold-down screen was attached to the waistcoat that held the wearable. The glasses-mounted display used was the MicroOptical SV-9 monocular display. In both cases, the display had a resolution of 640x480. The experimental task required the participants to walk between six locations in a lab while monitoring both the display of the wearable computer and their environment. As with the previous experiment, images were projected onto the walls of the lab with the participants required to acknowledge hazards. In this case, however, only images of the same form as those used in the image condition of the previous experiment were used. The participants were required to acknowledge the presence of a hazard simply by shouting "car," whereas in the previous experiment the participants were required to press a button on the experimental interface. This form of acknowledgement was chosen so as to remove a confounding variable—the participants' use of an unfamiliar interaction device: a handheld mouse. Acknowledging the hazards this way meant that once the participants had signalled they were ready they could move the cursor over the only active button on the experimental interface and press it when required without needing to move the mouse again. When a participant shouted "car" the experimenter recorded this acknowledgment using an experimenter application running on an iPaq. The experimental interface is shown in Figure 4. While moving between locations in the lab, the participants were also required to monitor the three characters displayed in the centre of the interface. When the three characters matched, the participants acknowledged this by pressing the "Match" button on the interface.

The layout of the lab is shown in Figure 5. The participants were required to walk from their current location to the location indicated in the top left corner of the experimental interface. The locations were physically indicated by squares approximately a meter across marked on the lab floor with the location number marked both in the centre of the square and behind the location on the wall. When a participant placed both feet in the square the next location was updated by the experimenter using the experimenter application. As with the acknowledgements of the hazards, this technique eliminated the need for the participant to move the cursor to press a different button—eliminating a confounding variable. At any given time the two projectors behind the location to which the participants were walking were used to display potential hazards. The definition

Figure 4. The experimental interface showing the next location to move to (top left), the characters to be matched (centre), and the "Match" button (centre left) which is pressed when the characters match. The "Ready" button (bottom left) is used only to signal the participants' readiness to start and is subsequently disabled. Both buttons flashed red when pressed to confirm that a press occurred.



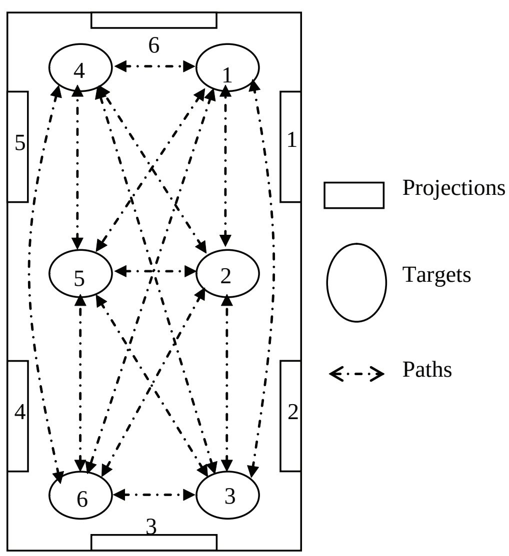
of ‘behind’ depended not only on the destination but also the previous location. For example, if a participant was walking from location 5 to location 2, projectors 1 and 2 would be used. If, on the other hand, a participant was walking from location 3 to location 2, projectors 1 and 6 would be used.

Experimental Procedure

The experiment consisted of two conditions run between three groups. The conditions were display: glasses mounted (GMD) or fold-down screen (FDS). In the GMD condition, if the participant wore glasses they could chose to clip the display onto their own glasses or wear the glasses that came with the display over their own glasses. The FDS was not removed from the waistcoat containing the wearable but the screen was folded up so it was not visible. In the FDS condition, the participants were required to wear glasses (either their own or the non-prescription glasses that came with the GMD) but without the display attached to remove glasses as a confounding variable. The groups were hazard distractions: none (No Distraction); single set of images (Single Distraction); or multiple sets of images (Multiple Distraction). The hazards were created by mapping one of seven characters to images. In the single distraction group, each character mapped to a single image. In the multiple distraction group, each character mapped to three images. In both cases only one character represents a hazard. A total of 18 participants performed the experimental task under both conditions; with the participants split equally across the three groups. The order in which the conditions were undertaken was counter-balanced within each group to eliminate any training effect.

Each condition consisted of a short training session where the participants were able to familiarize themselves with the task to be undertaken and the particular display being used. This training task required the participants to move to 18 different locations while monitoring their display and environment. After the training task, the experimental task—which consisted of moving between 60 locations in the lab—was undertaken. The paths the participants were required to follow were designed so that in each full condition the participant had to walk between each location in each direction exactly twice. To eliminate the

Figure 5. Lab layout showing the different locations participants were required to walk between and the location of the projections used to present potential hazards



paths as an experimental variable, every participant was required to navigate the same paths in the same order regardless of condition. Similarly, the same projection sequences were used in the same order for each participant. After completing the condition, the participants were required to complete a NASA TLX workload questionnaire (Hart & Staveland, 1988) to give an indication of the subjective workload experienced.

Results

The most surprising result was that a series of two-factor ANOVA tests showed no significant differences in either the subjective or quantitative results between the two displays. Total task time ($F_{1,30}$=0.00, p=0.992), percentage missed hazards ($F_{1,30}$=0.02, p=0.895), percentage missed matches ($F_{1,30}$=0.34, p=0.564), and overall workload ($F_{1,30}$=0.80, p=0.379) were all statistically similar.

Two-factor ANOVA tests did indicate, however, that the different distraction groups did have an effect on the results, with significant differences found in total task time ($F_{1,30}$=0.4.25, p=0.024), percentage missed matches ($F_{1,30}$=4.04, p=0.028), mental demand ($F_{1,30}$=5.18, p=0.012), physical demand ($F_{1,30}$=4.65, p-0.017), effort expended ($F_{1,30}$=3.89, p=0.032), performance level achieved ($F_{1,30}$=12.72, p<0.001), and overall workload experienced ($F_{1,30}$=6.60, p=0.004). *Post Hoc* Tukey HSD showed that, as one would expect, the requirement to monitor the environment had a negative effect on the results. The average total task time for the no distraction group (535.23 seconds) was significantly faster than for the single distraction group (618.85 seconds, p=0.0248) but not for the multiple distraction group (553.6 seconds, p=0.0938). The percentage of missed character matches was again significantly lower in the no distraction group (1.01%) compared to the single distraction group (7.61%, p=0.0403) but not the multiple distraction group (7.04%, p=0.0651). Similarly, the average mental demand experienced by the participants was significantly lower in the no distraction group (10.25) compared to the single distraction group (14.25, p=0.0121) but not the multiple image group (13.33, p=0.0616); and the average physical demand in the no distraction group (3.83) was significantly lower compared to the single distraction group (9, p=0.015) but not the multiple distraction group (7.33, p=0.124). The effort expended by the participants, however, was significantly lower in the no distraction group (8.33) when compared to the multiple distraction group (12, p=0.0418) but not the single distraction group (11.58, p=0.0778). The performance level the participants felt they achieved was significantly higher in the no distraction group (15.5) than both the single distraction (12.83, p=0.0369) and multiple distraction (10.33, p<0.001) groups. The overall workload experienced by the participants was also significantly lower in the no distraction group (6.98) than in the single distraction (9.84, p=0.013) and multiple distraction (10.028, p=0.008) groups.

Discussion

The most surprising aspect of the results was that the participants performed equally as well with the fold-down screen as with the glasses-mounted display. At first glance, it is reasonable to think that the participants would perform better when using the GMD as both the characters to be matched and the hazards to be noticed would be in the same field of vision. This, however, was not the case with participants missing fewer hazards in the FDS condition (8.48%) than in the GMD condition (8.96%) although this difference was not significant. Participants did miss fewer character matches in the GMD condition (4.61%) than in the FDS condition (5.84%) but, again, this difference was not significant. Even the physical demand experienced by the participants showed no significant differences ($F_{1,30}$=1.05, p=0.315) between the GMD (6) and FDS (7.44) conditions. This was surprising given the need to constantly bend down and look at the fold-down screen. Furthermore, due to the way the FDS was mounted (essentially on a belt), the screen had a tendency to angle away from the participants, requiring them to hold it in place. These results may, however, be explained by the difficulty participants experienced focusing on the

Table 3. Summary of the results highlighting the impact of monitoring the environment on the participant's task performance

	Avg. Task Time (Secs.)	Avg. % Of Missed Matches	Avg. Overall Workload
Single Distraction	618.85	7.61%	9.85
Multiple Distraction	553.61	7.04%	10.03
No Distractions	535.24	1.02%	6.99

GMD. Although the GMD was in the same field of vision as the projected images, the participants still found it necessary to actively shift their focus from their surroundings to the screen. This may be explained in terms of binocular rivalry (or, simply put, the confusion caused by two eyes seeing different things) and the different depth of focus required by the two eyes as discussed by Laramee & Ware (2002).

It was not surprising, however, that requiring the participants to monitor their environment while moving around the lab had a detrimental effect on their performance (Table 3). What is interesting, however, is that there were no significant differences in task performance between the single and multiple distraction groups. The single distraction group, on average, took slightly longer to complete the task and missed slightly more character matches but neither of these differences was significant. The multiple distraction group, on the other hand, experienced a slightly higher overall workload but, again, this difference was not significant. What this would seem to imply is that the participants were required to devote their attention to the two main tasks—monitoring the characters on the display and monitoring the projections for hazards—and the participants devoted slightly different amounts of effort to each. The overall performance, however, was approximately equivalent.

DISCUSSION

This chapter has described two evaluations of mobile systems that attempted to increase the realism of the lab environment through the use of visual distractions. In the first experiment, the participants' use of the audio cues was noticeably lower in the audio-visual condition when there were no visual distractions. This was most likely due to the participants' visual focus not being required anywhere other than on the device's screen. When the participants' visual focus was required elsewhere (to monitor the projections for hazards), the participants were prepared to make the added cognitive effort to parse the sounds. In the second evaluation, the % of character matches missed by the participants was significantly greater when the participants were required to monitor the environment for hazards. What these results demonstrate is that when evaluating a mobile application it is important to consider more than just mobility when considering the context in which to undertake the evaluation.

While it is not possible to draw any conclusions regarding the realism of the environment created, it would appear that the use of pseudo-realistic distractions is preferable to more abstract distractions. In the first evaluation, participants were better able to discern the hazards when presented as characters as opposed to photographic images. While the use of different forms of abstract distraction may reduce this problem, the use of images would seem to be the simplest approach. Interestingly, the use of a single set of images (with one hazard image out of a total of seven) did not have any effect on the results when compared to the use of multiple sets of images (with 3 hazard images out of a total of 21). This implies that the participants were not memorizing the hazards, meaning that if the distractions are of an appropriate form, a relatively small set is appropriate. Future work is necessary to validate these hypotheses and to determine whether the form of distractions presented here provide results that are similar to a real-world scenario.

It is also worth mentioning the importance of experimenter mobility in the two evaluations presented. In both instances, a mobile application allowed the experimenter to interact with the experimental server. This eased the running of the evaluations by allowing the experimenter to move around the lab, assisting the participant when necessary. For example, in both cases, the experimenter could ensure that the participant was in the appropriate location at the start of the experiment and was ready to begin before starting the experiment. While this could have been achieved from the observation room off the lab, the mobility of the experimenter made this easier. It also enabled the participants to acknowledge the hazards in the second evaluation without having to shout to an unseen experimenter in the observation room.

CONCLUSION

This chapter has described two evaluations which demonstrated that the use of visual distractions can have a significant effect on the results of a usability evaluation of a mobile device. Participants' task performance was significantly affected by the requirement to monitor their environment for such distractions and the subjective workload experienced was also increased. Interestingly, there was some evidence that the introduction of such distractions changed the results of an evaluation, with the workload experienced when using audio navigational cues—compared to visual cues—being significantly higher when no distractions were present but being the same when distractions were used. This would indicate that—although further work is required in this field—the use of visual distractions is something that should be considered when designing an evaluation of a mobile application.

ACKNOWLEDGMENT

We would like to thank all the experimental participants for their participation. The software was written in the EWE (www.ewesoft.com) and Java (www.java.com) programming languages. The sounds were designed using MIDI Studio V4.20 (www.sonicspot.com/midistudio/midistudio.html) and converted to wave files for playback on an iPaq using Midi2Wave Recorder V3.5 (www.midi2wav.com). This work was performed under NRC REB ethics approval numbers 2005-46 and 2005-47.

REFERENCES

Blattner, M. M., Sumikawa, D. A., & Greenberg, R. M. (1989). Earcons and icons: Their structure and common design principles. *Human-Computer Interaction, 4*(1), 11-44.

Brewster, S. (2002). Overcoming the lack of screen space on mobile computers. *Personal and Ubiquitous Computing, 6*(3), 188-205.

Crossan, A., Murray-Smith, R., Brewster, S., Kelly, J., & Musizza, B. (2005). Gait phase effects in mobile interaction. In *CHI '05 Extended Abstracts on Human factors in Computing Systems* (pp. 1312-1315). Portland, OR, USA: ACM Press.

Duh, H. B.-L., Tan, G. C. B., & Chen, V. H.-h. (2006). Usability evaluation for mobile device: a comparison of laboratory and field tests. In *Proceedings of the 8th Conference on Human-computer Interaction with Mobile Devices and Services (Mobile HCI 2006)* (pp. 181-186). Helsinki, Finland: ACM Press.

Goodman, J., Brewster, S. A., & Gray, P. D. (2004). Using field experiments to evaluate mobile guides. In *Proceedings of HCI in Mobile Guides (Workshop at MobileHCI 2004)*, Glasgow, Scotland.

Hart, S., & Staveland, L. (1988). Development of NASA_TLX (Task Load Index): Results of empirical and theoretical research. In P. Hancock & N. Meshkati (Eds.), *Human mental workload* (pp. 139-183). Amsterdam, Holland.

Holland, S., Morse, D. R., & Gedenryd, H. (2002). AudioGPS: Spatial audio navigation with a minimal attention interface. *Personal and Ubiquitous Computing*, *6*(4), 253-259.

Kjeldskov, J., Skov, M. B., Als, B. S., & Høegh, R. T. (2004). Is it worth the hassle? Exploring the added value of evaluating the usability of context-aware mobile systems in the field. In S. Brewster & M. Dunlop (Eds.), *Mobile human-computer interaction - Mobile HCI 2004* (Vol. 1, pp. 71-73). Glasgow, Scotland: Springer-Verlag.

Kjeldskov, J., & Stage, J. (2004). New techniques for usability evaluation of mobile systems. *International Journal of Human Computer Studies (IJHCS)*, *60*, 599-620.

Laramee, R. S., & Ware, C. (2002). Rivalry and interference with a head-mounted display. *ACM Transactions Computer-Human Interaction*, *9*(3), 238-251.

Loomis, J. M., Marston, J. R., Golledge, R. G., & Klatzky, R. L. (2005). Personal guidance system for people with visual impairment: A comparison of spatial displays for route guidance. *Journal of Visual Impairment & Blindness*, *99*(4), 219-232.

Lumsden, J., Brewster, S., Crease, M., & Gray, P. D. (2002). Guidelines for audio-enhancement of graphical user interface widgets. In F. Détienne, X. Faulkner & J. Finlay (Eds.), *Proceedings of HCI 2002* (Vol. II, pp. 6-9). London: Springer-Verlag.

Mizobuchi, S., Chignell, M., & Newton, D. (2005). Mobile text entry: relationship between walking speed and text input task difficulty. In *Proceedings of the 7th international Conference on Human Computer Interaction with Mobile Devices & Services* (pp. 122-128). Salzburg, Austria: ACM Press.

Mustonen, T., Olkkonen, M., & Hakkinen, J. (2004). Examining mobile phone text legibility while walking. In *CHI '04 Extended Abstracts on Human Factors in Computing Systems* (pp. 1243-1246). Vienna: ACM Press.

Öquist, G., & Goldstein, M. (2001). Towards an improved readability on mobile devices: Evaluating adaptive rapid serial visual presentation. In F. Paternò (Ed.), *Proceedings of Mobile HCI 2002* (pp. 225-240). Pisa, Italy: Springer-Verlag, Berlin Heidelberg.

Petrie, H., Johnson, V., Strothotte, T., Raab, A., Fritz, S., & Michel, R. (1996). MOBIC: Designing a travel aid for blind and elderly people. *Journal of Navigation*, *49*(1), 45-53.

Revels, A., & Kancler, D. (2000). Evaluation of mobile computing displays. In *Proceedings of SPIE*, *2000* (Vol. 4021, pp. 33-44).

Sheedy, J., & Bergstrom, N. (2002). Performance and comfort on near-eye computer displays. *Optometry and Vision Science*, *79*(5), 306-312.

Starner, T. (2003). The enigmatic display. *Pervasive Computing*, *3*(1), 15-18.

Strachan, S., Eslambolchilar, P., Murray-Smith, R., Hughes, S., & O'Modhrain, S. (2005). GpsTunes: controlling navigation via audio feedback. In *Proceedings of the 7th International Conference on Human Computer Interaction with Mobile Devices & Services* (pp. 275-278). Salzburg, Austria: ACM Press.

KEY TERMS

Active Distraction: A distraction that a user must respond to in some way. Examples of such distractions include hazards that a user must avoid (e.g., a lamp-post) or a mobile-phone that the user answers when it rings.

Earcon: Abstract, structured sounds used to provide information to a user. The musical qualities of the sound (e.g., rhythm, timbre, or pitch) can be varied to convey information to users. Earcons can be combined sequentially (compound earcons) or concurrently (parallel earcons).

Field Studies: The evaluation of a mobile application that takes place in the actual context of use. The advantage of such evaluations is that problems that only arise in the particular context will be detected. The disadvantages of such evaluations include difficulties in controlling the environment and capturing evaluation data.

Head-Mounted Display: A display that allows a user to view the visual output of a wearable computer at all times. Such displays may cover one eye (monocular) or two (binocular). They may be opaque (for immersive environments) or transparent (allowing the user to simultaneously view the surrounding environment). The displays may attach to a pair of glasses (glasses-mounted display) or be attached to a form of headwear such as a hat or band.

Interfering Distraction: A distraction that interferes with a user's ability to interact with their mobile device. Such distractions may be passive or active. Examples of such distractions include traffic noise interfering with a mobile device's audio feedback or oncoming pedestrians that limit a user's ability to monitor the visual display of a mobile device.

Lab Evaluations: The evaluation of a mobile application that takes place in a laboratory. The advantages of such evaluations include ease of controlling the environment and data capture. Disadvantages include difficulties in creating an appropriately realistic evaluation setting.

Passive Distraction: A distraction that may "put-off" the user but can be ignored. Examples of such distractions include billboards mounted on the side of buildings or the sound of traffic on a distant road.

Wearable Computer: A mobile computing device that the user wears rather than carries. The processing unit may be worn in a pouch on a belt or in a bag. Various interaction devices may be secreted around the user's body including: a handheld input device such as a trackball, a wrist mounted keyboard, or a head-mounted display. The advantages of such devices over handheld mobile devices include: faster access to the device, increased privacy, and hands-free access to data.

ENDNOTE

1 C3 is middle C (261.63Hz)

Chapter LV
Instrumented Usability Analysis for Mobile Devices

Andrew Crossan
University of Glasgow, UK

Roderick Murray-Smith
University of Glasgow, UK
Hamilton Institute, NUI Maynooth, Ireland

Stephen Brewster
University of Glasgow, UK

Bojan Musizza
Institut Jozef Stefan, Slovenia

ABSTRACT

Instrumented usability analysis involves the use of sensors during a usability study which provide observations from which the evaluator can infer details of the context of use, specific activities, or disturbances. This is particularly useful for the evaluation of mobile and wearable devices which are currently difficult to test realistically without constraining users in unnatural ways. To illustrate the benefits of such an approach, a study of touch-screen selection of on-screen targets is presented whilst walking and sitting, using a PocketPC instrumented with an accelerometer. From the accelerometer data, the user's gait behaviour is inferred, allowing linking performance to gait phase angle, showing there were phase regions with significantly lower error and variability. The chapter provides examples of how information acquired via sensors gives quantitatively measurable information about the detailed interactions taking place when mobile, allowing designers to test and revise design decisions, based on realistic user activity.

INTRODUCTION

Mobile and wearable devices are becoming increasingly important in our daily lives, and there is a correspondingly large activity in the design of interaction for these devices. It is obviously very important to be able to evaluate their usability, but by their very nature, these devices are intended for use in mobile settings, not for use by someone seated in a usability lab.

As described by Kjeldskov and Stage (2004), there is a wealth of guidelines for running laboratory-based usability studies, but these studies will lack realism for mobile devices. To test mobile devices in mobile settings, however, it is required to use field-based evaluations, which are far from straightforward to implement. Kjeldskov and Stage's review of the literature points out three difficulties: 1. It is difficult to define a study that captures the use-scenario, 2. It is hard to use many established evaluation techniques, and 3. Field evaluations complicate data collection and limit experimental control. Examples of papers where researchers have proposed additional techniques such as distance walked and percentage preferred walking speed to assess usability include Brewster (2002); Petrie, Furner, and Strothotte (1998); and Pirhonen, Brewster, and Holguin (2002), using a mix of qualitative questions and manual recording of walking pace. Mizobuchi, Chignell, and Newton (2005) examine the effect of key size on handheld devices while walking.

Barnard, Yi, Jacko, and Sears (2005) review the differences between desktop and mobile computing, and they observe "for researchers aiming to isolate the effects of motion from other contaminants, the idea of such uncontrolled studies can be daunting. Control is critical for empirical data collection methods employing the scientific method." Roto et al. (2004) discuss the use of Quasi-experimentation based on best possible control over nuisance variables, coupled with recordings of the user, interaction with the device, and environment. The innovation in their recordings was the use of multiple cameras worn around the body of the user and attached above the screen of the mobile device. This does make the recording process obtrusive and might change both user behaviour and that of people in the environment around them. It is also time-consuming to analyse after the experiment. This recording arrangement has been used successfully in Oulasvirta, Tamminen, Roto, and Kuorelahti,(2005) to investigate the fragmentation of attention in mobile interaction.

INSTRUMENTED USABILITY ANALYSIS

Here, '*Instrumented usability analysis*' is defined as the use of sensors during a usability study which provide observations from which the evaluator can infer details of the context of use, or specific activities or disturbances.

Sensors such as accelerometers, magnetometers, and GPS systems have been added to mobile devices and are now in mass-production in mobile phones. These have been included for informing the user (about location, number of steps taken) or giving the user novel input mechanisms, such as gesture recognition or input for game playing.

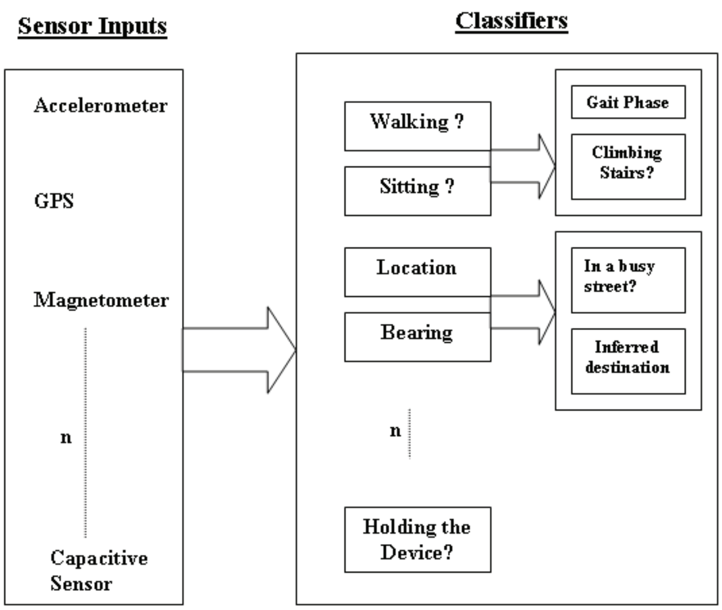
There are many examples of both prototype and commercially available sensors and sensor packs for motion or context sensing. Fishkin, Jiang, Philipose, and Roy (2004) describe a system for detecting interactions with RFID technology and suggest it can be used to infer user movement by examining signal strengths from a sensor network. Gemmell, Williams, Wood, Lueder, and Bell (2004) describe the SenseCam system used to capture life experiences without having to operate complex recording equipment. SenseCam combines a camera with a group of sensors including an accelerometer, infrared, light and temperature sensors, and a clock to automatically detect, photograph, and map out changes in context or events during a persons day. Kern and Schiele (2003) describe a hardware platform combining multiple wearable accelerometers in order to infer the user's context and actions. They demonstrate how these acceleration signals can be used to classify user activity into actions such as sitting, standing, walking, shaking hands, and typing.

More recently, the general purpose Bluetooth Sensing Hardware Accessory for Kinesthetic Expression (SHAKE) inertial sensor pack, described in Williamson, Murray-Smith, and Hughes (2007), which is available for general use by the research community, from SAMH Engineering Services. It features a tri-axis accelerometer, tri-axis magnetometer, dual channel analogue inputs, dual channel capacitive sensing, and an internal vibrating motor. Communications are over a Bluetooth serial port profile. SHAKE includes a powerful DSP engine, allowing real time linear phase sample rate conversion. These capabilities allow rapid prototyping of inertial-sensing-based interfaces with real-world hardware. But the small size, and the onboard processor and memory mean that the device can be used completely separately from the implementation on a mobile device, and can be used to log movement at multiple points around a user in a wide variety of situations. It can be attached to the back of their device, or could be attached to their belt, or elsewhere on the body, to detect activity without wires to restrict a person's movements.

In this chapter it is suggested that such sensors and sensors packs can be used in an indirect fashion, to better understand what was happening to the device and user at any point in time during a usability experiment. Figure 1 demonstrates how such a system would work. There could potentially be multiple sensors placed on the user or mobile device. Outputs from the sensor would be run through one or many classifier algorithms that could infer the user's actions or context at any one time.

Of course, as the algorithms for automatically inferring context of use from sensors develop to a level of robustness which allows them to be used online, they can be used in everyday mobile situations to subtly adjust the nature of the interaction—the "background interaction ... using naturally occurring user activity as an input that

Figure 1. An example of how sensors and classifiers could be combined to infer user actions during a usability study and store these as an annotated log file, allowing developers to correlate different states with user interaction behaviour. Raw readings are interpreted in a hierarchical fashion by a range of plug-in classification or signal transformations. These can be arranged hierarchically so, for example, only if the person is classified as "walking," is gait phase angle inferred, and whether they are climbing the stairs or not.



allows the device to infer or anticipate user needs" described in Hinckley, Pierce, Horvitz, and Sinclair (2005). For example, if tapping is less accurate when a user is walking, the display could adjust to a mode which had fewer, but larger, buttons. In this chapter, however, the use of such information to analyse user behaviour in greater temporal detail than is typical in mobile usability trials will be concentrated on. Rather than performing a trial and then asking subjective questions or analyzing video footage, activity from sensors on the device or user will be classified, and these will be related to any log of explicit interaction activity during the evaluation. If a user has an unusually high error rate, can it be better determined exactly what was happening at that point in time in each case?

This approach is obviously related to research in context-dependent interaction which used information from sensors to infer context of use (walking, running, in car, inside, outside), to allow more appropriate behaviour from the device. Yi, Choi, Jacko, and Sears (2005) used an accelerometer in evaluation, but used mean activity over the whole period in different conditions, rather than detailed results during the evaluation.

These techniques can be combined with a system such as Replayer described by Morrison *et al.* (2006). This is a system designed to aid usability evaluation and provides tools that allow evaluators from different backgrounds to easily view, annotate, and analyse multiple streams of heterogeneous data logged during a usability study. These data streams could potentially be obtained from multiple sensors attached to a mobile device. A first prototype of this, using the MESH device used in this chapter, is described in Morrison, Tennent, and Chalmers (2007).

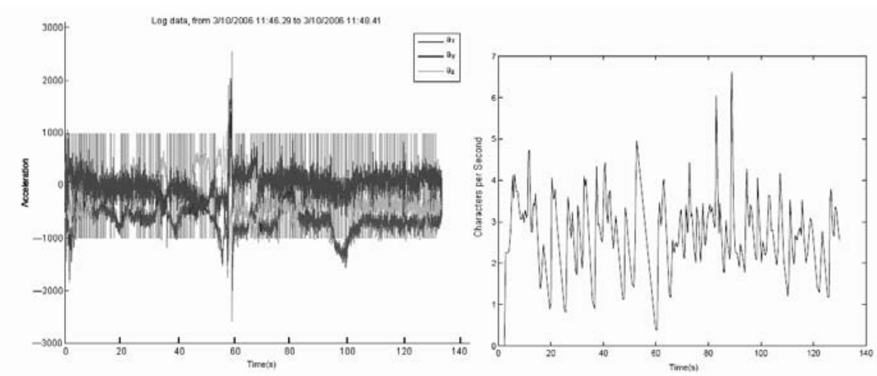
Example: Mobile Text Entry

An illustration of how the method could be used is that of mobile text entry. The questions that may want to be answered for a given method could be: Do people use it on the go, as well as when stationary? How much slower are they when they use it while walking? Do they enter text continuously, but slowly, or do they stop every few metres to enter more text? How is their error rate related to their walking speed? Do they link the entry of a new character with a new step? What is the effect on walking speed, when entering text? If the user enters text in a car or bus, how are they affected by movement of the vehicle? Do they wait until the bus stops then enter text?

Figure 2 shows a time-series of accelerometer readings while a user enters text, while seated as a passenger in a car in urban rush hour traffic with frequent stop-start activity.

Figure 3 shows the example of a user entering text in various contexts. The user started to enter text while sitting, standing up, walking around

Figure 2. Text entry while seated in a car. The number of characters entered per second is plotted to the right of the acceleration time series (the vertical lines in the left plot are individual key-press events).



some narrow corridors avoiding objects, down stairs, along a straight corridor, up stairs, then returned to a seated position, and entered more text while resting his hand on a table. The plots show the overall activity of the user along with the throughput of characters entered at each point. In the walking case, text entry pause can be seen as the user takes a seat just after 140s, and faster entry rates while seated in the car, compared to walking are seen. The text entry rates while walking are nevertheless fairly constant. This illustration acts as an example of how the accelerometers can give us extra information from which we can infer more about what was happening at each point in the interaction.

In this chapter, which expands on earlier work in Crossan, Murray-Smith, Brewster, Kelly, and Musizza (2005), a quantitative understanding of the detailed interactions taking place is worked towards via additional sensors on the mobile device and user, so that it can be better understood how users interact with the devices, and so further improve the designs.

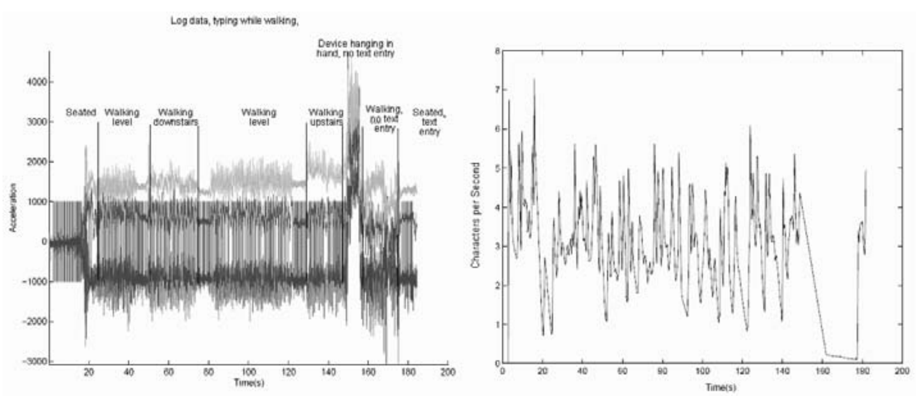
DETAILED CASE STUDY OF WALKING AND TAPPING

Here, a detailed example of an instrumented usability study to demonstrate the benefits of this approach is presented. Standard usability time and error metrics are gathered, while the instrumentation allows gaining a greater insight into the users actions and disturbances during the study. Although this example specifically examines users walking patterns sensed through an accelerometer, the techniques discussed can be applied to a wide range of contexts.

Given the importance of devices being used while the user is walking, and the difficulty researchers have had about getting detailed insight into user behaviour, down to the level of each step taken, the example of tapping buttons or other widgets on a touch screen is now concentrated on. This is a common form of input, and is effective when seated, but difficult when walking. Brewster (2002) showed a more than 30% reduction in performance, tapping buttons on the display of a PDA when walking, compared to sitting. If more detailed insight into how and when users tap during walking can be gained, the design of the interface to improve robustness may be able to be adjusted.

Here, it will be shown how sensors like accelerometers can be used in ways other than for explicit interaction. In this case, the acceleration data to infer the user's gait is used, and whether the rhythm of walking affects the tap timing and error rate of a user selecting targets on screen, while walking and sitting is investigated.

Figure 3. Text entry in a range of conditions. The number of characters entered per second is plotted to the right of the acceleration time series (the purple vertical lines in the left plot are individual key-press events, and the black ones delineate the different walking conditions).



Experiment

Introduction

This study examines in detail the behaviour of users tapping on the screen of a mobile device. It analyses behaviour in two different situations that a user might perform this task: while sitting and while walking. The sitting condition will be used to provide a performance baseline for the walking condition. Disturbances to the device and the users stylus due to the user's walking, will affect how and how well the user performs the task. By instrumenting the device with a sensor (in this case an accelerometer), a deeper understanding of how these disturbances affect performance, will be able to be gained.

Equipment

This system was developed using an HP 5550 PDA with the Xsens P3C 3 degree of freedom linear accelerometer attached to the serial port, as shown in Figure 4. Its effect on the balance of the device is negligible (its weight is 10.35g). The accelerometer was used to detect movement of the device, sampling at a rate of approximately 90Hz.

Figure 4. PDA with the Xsens P3C accelerometer attached to the serial port

Task

The interface used for the study is displayed in Figure 4. Participants were asked to tap on a series of cross-hair targets (drawn 30 pixels high and wide) that were displayed on the screen. There were 15 possible target positions spaced equally around a 3 wide by 5 high grid of positions on the screen. Every second target presented to the participants was the target in the centre of the screen. This ensured that the user must return the stylus to the centre of the screen such that when a target other than the central target was tapped, the path to that target was always from the centre. The other 14 targets were displayed to the user in a random order four times each. The accuracy and speed of tapping were both emphasised as equally important. The position of the tap was recorded as the initial stylus down position on the screen. Once one target had been selected, the next target was displayed a random time interval from 0.5 to 1.5 seconds after the previous selection. This was to prevent rhythm effects affecting the tapping phase information in the mobile condition. There were no restrictions on the accuracy that was required by the user. A tap anywhere on the screen regardless of the position of the target counted as a selection.

There were two experimental conditions: tapping while sitting and tapping while walking, and 20 users performed both conditions in a counterbalanced order, with 18 participants being right handed and 2 participants being left handed. All participants tapped with their dominant hand while holding the device in their non-dominant hand. For the walking condition, the participants navigated a quiet triangle of paths on the university campus (of total length, approximately 200 metres).

Calibration of the screen becomes an issue when looking at accuracy of tapping in a pen based interface, as an error in the calibration can lead to a consistent and unwanted bias in the results. The screen was calibrated once at the start of the experiment, and the same device was used throughout the experiment. Three participants tested the screen calibration. The device was placed on the desk and users performed a similar task to the tapping study for four separate sessions. In this case, accuracy was heavily emphasised as the most important aspect of the study. This was borne out by the much closer concentration of points than in the final results, with mean standard deviation of the error for each participant for all targets being less than a pixel. After each session, the device was rotated by 90 degrees (additive for each section) to negate any systematic tapping bias. Mean values were recorded for each screen target position and were subtracted from the final results. This method provides a closer match between the position the user actually tapped in and the recorded tap position.

Metrics

Standard usability metrics were used for assessing user performance in the task. Comparisons were made between time to tap and accuracy of tap for each of the groups. Time to tap was taken from the time that the target was displayed on the screen to the time of the stylus down event. The hypotheses were that users would be more accurate and faster in the seated condition. The effect of screen position of the target on accuracy of the tap was also examined.

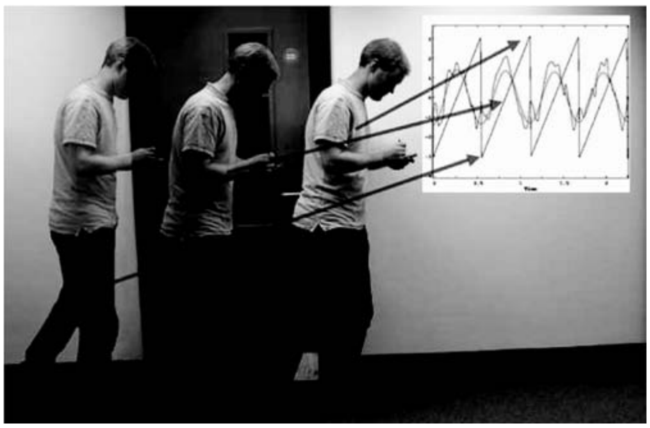
The instrumented usability approach also allows gaining further insights into the users actions during the study. The interactions of participants' tapping and step patterns were examined.

Gait Detection

As a mobile user walks while holding a mobile device, his or her arm will oscillate as a result of the user's gait. If only the vertical axis of this oscillation is examined, there will be one oscillation per step. Figure 5 shows a time series for the vertical acceleration axis. A Fast Fourier Transform is used to determine the frequency at which the peak amplitude occurs, between 1 and 3Hz in the spectrum. For the controlled conditions in this study, this corresponds to the walking step rate. In practice, this is the frequency of maximum power in the spectrum as the users are looking at the screen and therefore trying to hold the device relatively still with respect to their body as they walk. The vertical axis acceleration signal is then zero phase shift, filtered using a narrow bandpass Butterworth filter centred around this frequency. Figure 5 demonstrates the filtered signal. As the user walks with the device held steady in one hand, an approximately regular oscillation is formed in the vertical axis. One oscillation corresponds to one step.

The algorithms used in this chapter were developed in research on synchronisation effects in

Figure 5. A user walking with the device and corresponding acceleration trace. The unfiltered vertical acceleration signal (rough sinusoid), the filtered signal (smooth sinusoid), and the phase estimate (in radians) for the signal (saw-tooth).



nature. The oscillations involved in many natural systems are often irregular, ruling out simple strategies. In some cases, such as respiratory examples, or electrocardiogram data, there are clear marked events with pronounced peaks in the time-series which can be manually annotated or automatically detected. One practical advantage of the use of the synchronization theory is that often we have a quite complex nonlinear oscillation, which might be sensed via a large number of sensors. The phase angle of that oscillation is however a simple scalar value, so if the synchronization effects in two complex systems is being investigated, the analysis can sometimes be a single value, the relative phase angle $\phi_1 - \phi_2$.

The Hilbert Transform

How is the phase angle from the data found? A common approach is to use the Hilbert transform introduced by Gabor in 1946, which gives the instantaneous phase and amplitude of a signal s(*t*) (Pikovsky, Rosenblum, & Kurths, 2001). The Hilbert transform signal $s_H(t)$ allows you to construct the complex signal:

$$\zeta(t) = s(t) + is_H(t) = A(t)e^{i\phi(t)}$$

where $\phi(t)$ is the phase at time *t*, and A(*t*) is the amplitude of the signal at time *t*. The Hilbert transform signal of s(*t*) is:

$$s_H(t) = \frac{1}{\pi} \lim_{T\to\infty} \int_{-T}^{T} \frac{s(\tau)}{t-\tau} d\tau$$

Although A(t) and $\phi(t)$ can be computed for an arbitrary s(*t*), they are only physically meaningful if s(*t*) is a narrow-band signal. For the gait analysis, therefore, the data is filtered to create a signal with a single main peak in the frequency spectrum around the typical walking pace (between 1 and 3Hz).

This phase plot signal is again shown as the saw-tooth waveform in Figure 5 and Figure 6 and can be seen to reset at the lowest point in the signal. This corresponds to the lowest point of the hand in the oscillation.

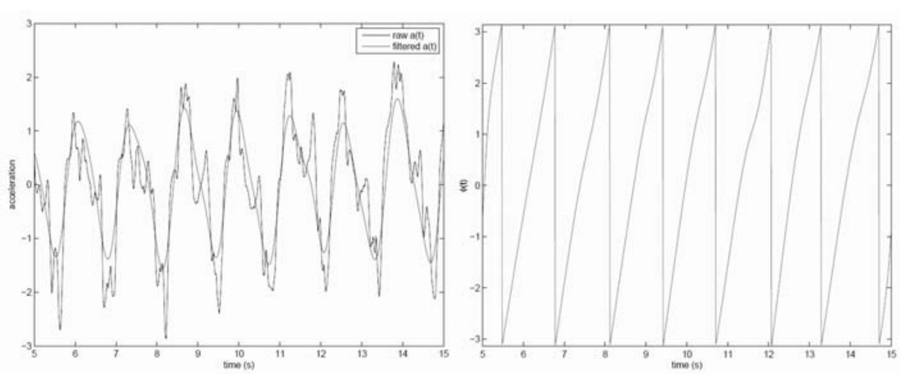
Details of the Hilbert transform and filtering are included here for completeness, however, this functionality is easily accessible in many standard data analysis programs such as Matlab, though simple function calls and understanding these equations is not essential for understanding the remainder of this chapter.

Standard Usability Results

Time to Tap

The mean time to tap was lower in the sitting case than the walking case, as would be expected. The mean time to tap a target in the walking condition was 0.79s (std dev = 0.18) compared to 0.70s (std

Figure 6. Generating the phase angle $\phi(t)$ from observed acceleration data a(t) from a user walking



dev = 0.22) in the seated case. This can be further broken down into tapping the centre target and outer targets. The mean time to tap the centre target was 0.75s (std dev = 0.23) when walking and 0.65s (std dev = 0.19) while sitting. This compared to 0.82s (std dev = 0.22) while walking and 0.75s (std dev = 0.20) while sitting to tap the outer targets. This difference between centre and outer targets is indicative of users predicting the appearance of the centre target since it consistently appeared every second target.

Tap Accuracy

A graph of tapping accuracy is shown in Figure 7. The graph demonstrates that as expected, users were more accurate tapping in the seated condition with 78% of taps being within 5 pixels in the seated case compared to 56.5% in the walking case. Participants remained more accurate in the seated case and reached 98% of taps within 15 pixels in the seated condition compared to 25 pixels in the walking condition. Separating these into *x* and *y* pixel error showed little difference between accuracy in vertical or horizontal error.

Above the range of 30 pixels, structure can be seen in the errors where tap position corresponds to the position of the previous target (shown in Figure 8). This indicates a tap when the user did not mean to tap. This is most likely the result of a user accidentally double tapping in position of the previous target. These taps were viewed as outliers and discounted from the final analysis.

Observation in the walking condition showed that when tapping, all participants immediately adopted the strategy of grounding the side of their hand holding the stylus on the hand holding the device to reduce independent movement of the hands and thereby improve accuracy. Targeting therefore involved pivoting the hand about the grounded position.

Figure 9 shows the mean variability and covariance of the *x* and *y* target errors for all users for each of the 15 targets. In almost all cases, the variability in tapping is smaller in the seated condition than in the walking condition. Due to the controlled conditions of this study, the movements to the outer targets were always from centre target. The variability in tap position for the centre targets is less than that of the outer targets. This is due to the fact that the stylus over the centre target position was the default position for most users. Covariance of the *x* and *y* tap positions can be seen to be along the direction of movement for most of the targets. This is particularly true for the corner targets.

Figure 7. Percentage of taps with the given pixel radius for sitting and walking users

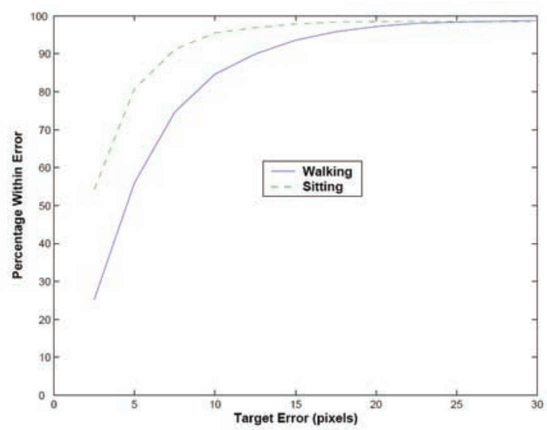


Figure 8. The x- y pixel errors for all users for all targets. The structure of the 3 by 5 grid of targets can be seen indicating users mistakenly double tapping

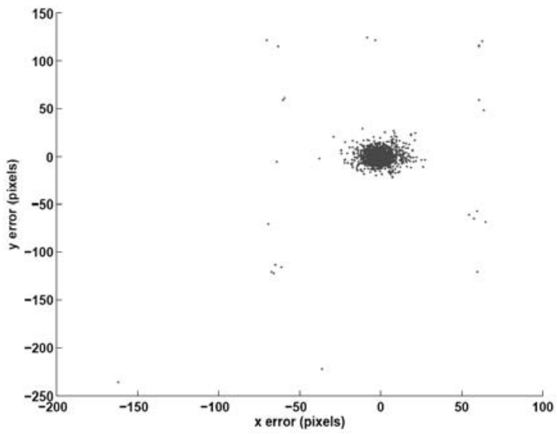


Figure 9. Ellipses show two standard deviations of a Gaussian fit to the spread of mean tap positions (from four points per participant) from all 20 participants, for each target. In each case the smaller ellipse shows the results for the seated condition and the larger ellipse shows the results for the walking condition. The crosses represent the target positions.

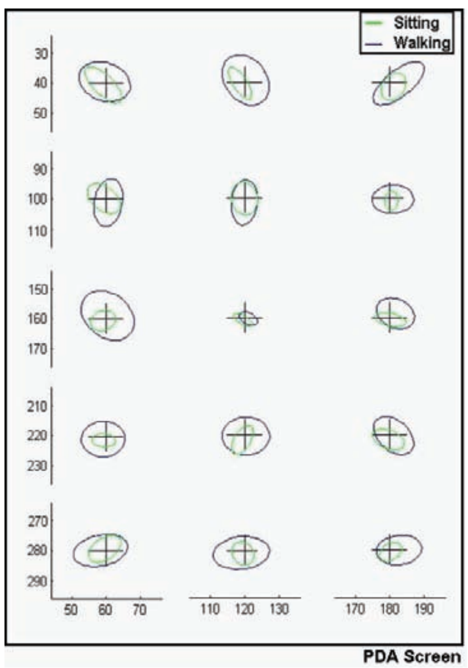
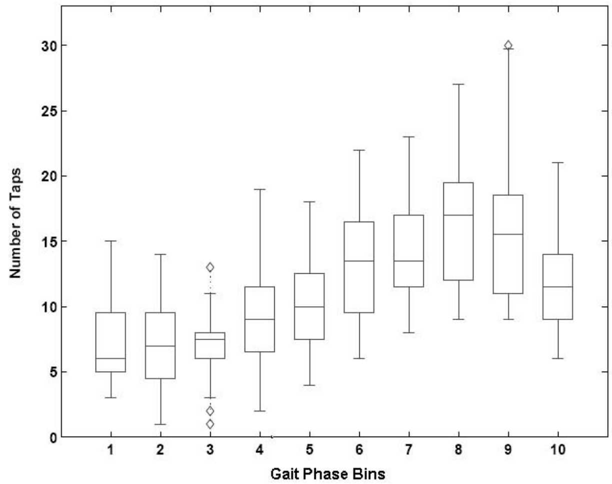


Figure 10. Box plot visualising the distribution of tapping times. Median phase in which the user taps (split into 10 sections) with the reset position for the phase corresponds to the lowest point of the arm which occurs just after a step.



Instrumented Usability Analysis Results

Tap Phase

The method for obtaining the phase of steps that the tap occurred at is described. Figure 10 splits one step into 10 equal sections and plots the median of the number of taps in each section for each participant. The reset phase position corresponds to the lowest point of the vertical accelerometer trace. Bins 1 to 5 correspond to the arm as it moves upwards to its peak, and bins 6 to 10 correspond to the arm moving downwards.

A bias in clearly shown towards tapping in the second half of the oscillation. This bias is not present when analysing the phase at which the targets are displayed and must therefore have been introduced by the user. The phases when most taps occur correspond to when the device is moving downwards with the arm. As soon as the device begins to move upwards in the hand again towards the stylus, the number of taps on the screen decreases. When questioned after the experiment, none of the participants was aware that a bias existed.

Figure 11 shows the median of the mean magnitude tap error for each participant, for each of the step phase bins. This figure shows that users were more accurate when tapping in the second half of the phase—the time when most taps occurred. The mean error is 7.1 pixels in the first section (just when the arm starts to rise again), compared to a mean of 5.6 pixels in the fourth section when the hand is moving downwards.

Further to this, if just the three most probable tap phase bins (P_{HP}) and the three least tap probable tap phase bins (P_{LP}) are considered, a clearer indication is given. Figure 12 shows a box plot of the tap error in P_{HP} and P_{LP}. P_{HP} has median tap error of 4.6 pixels compared to 5.7 pixels for P_{LP}. A Mann Whitney test showed that this difference was highly significant ($p < 0.002$). If timing data for the same phase regions is considered, it can be seen that users take significantly longer to tap in the high probability regions. Users took a median of 0.69 seconds to tap in P_{LP} compared to 0.73 seconds for P_{HP}. This difference was again tested using a Mann Whitney test and was shown to be significant ($p = 0.05$). Figure 13 shows the corresponding skew plot for high and low tap probability regions.

When combined with the results shown in Figure 10, these data suggest that users were able to subconsciously alter their behaviour in the task in order to improve their accuracy by tapping at a time in their step when it was easier

Figure 11. Median target tap error in pixels for each phase of the motion (split into 10 segments) with the reset position for the phase corresponding to the lowest point of the arm which occurs just after a step

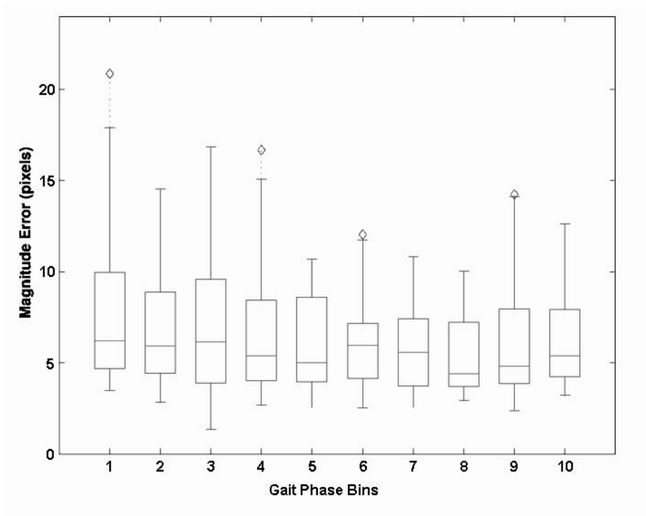
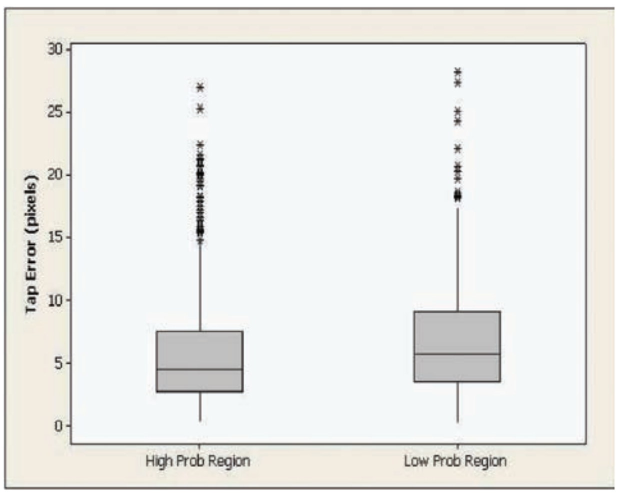


Figure 12. Target tap error in pixels for the high probability tap phase region and the low probability tap phase region

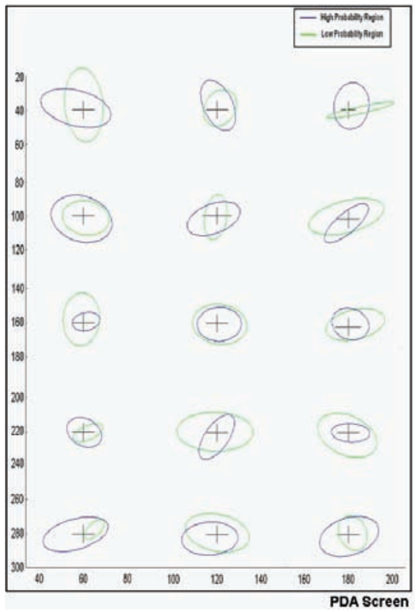


to tap more accurately. The longer time to tap in the high probability region indicates that users tended to subconsciously wait for that particular phase region to tap in.

Left–Right Step Analysis

For the previous set of results, each step has been treated as one cycle. However, it could also be chosen to separate out the left foot steps from the right foot steps. As the user walks, the vertical acceleration sensed through the device will complete one phase cycle at every step the user takes. The lateral acceleration can also be seen to be oscillatory. However, one oscillation in the lateral direction will now correspond to a combination of one left foot step and one right foot step. The dominant frequency of the lateral oscillation

Figure 13. Target tap error in pixels for the high probability tap phase region and the low probability tap phase region



is therefore half that of the vertical oscillation. The device will therefore undergo consistently different disturbances depending on whether the user is stepping with the left or right foot.

The data gathered from the accelerometer was analysed to separate left and right foot steps. The vertical acceleration was used to delineate the steps with the lateral acceleration used to determine right and left foot steps. Using this method ensures that valid comparisons can be made between the one step-one cycle data and the one step-two cycle data. Figure 14 shows the distribution of tapping through the phase of the left and right foot steps. The first eight bins correspond to a left foot step and the second eight bins to the right foot step. It can be seen from this figure that the tapping pattern for the two step per phase cycle data follows the one step per phase cycle data. There are distinct interactions visible between the tapping and the stepping in each step. The pattern for each step is consistent, with the peak tapping phase values occurring at around the foot down phase of the step for both left and right foot. There were no significant differences detected between the tap errors for left foot steps and right foot steps. The median magnitude of error for taps occurring during the left foot phase was 2.9 pixels compared to 2.8 pixels for taps during the right foot phase section. There were no significant differences with either the separated x error or the y error for taps during the left or right steps.

Walking Speed Analysis

Analysis of the participants' walking speed throughout the experiment showed that the step rate during the study was relatively consistent for all users. Figure 15 shows the estimated step rate for five typical participants over the duration of the study. The task chosen for the study was consistent throughout the task (tapping on a screen). The path that the participants traversed during the experiment was relatively quiet. There was therefore little reason for the participants to speed up or slow down their step rate. In other experiments where the path is more complex or the user must perform different tasks, it would be expected for the step rate to be more variable, but in this study, mean step rate might actually have been sufficient when analysing the effects of walking rate.

Figure 14. Box plot visualising the distribution of tapping times. Median phase in which the user taps (split into 16 sections). Unlike Figure 10, one phase cycle includes both a left and a right step.

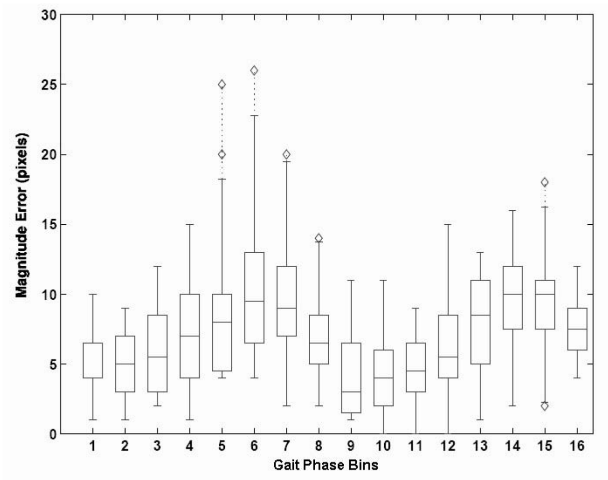


Figure 15. The step rate of five typical participants for duration of the walking condition

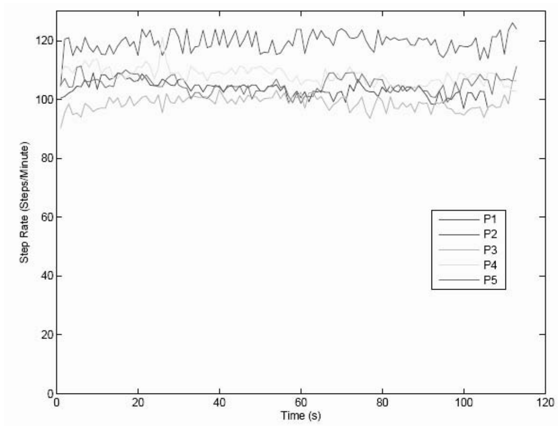
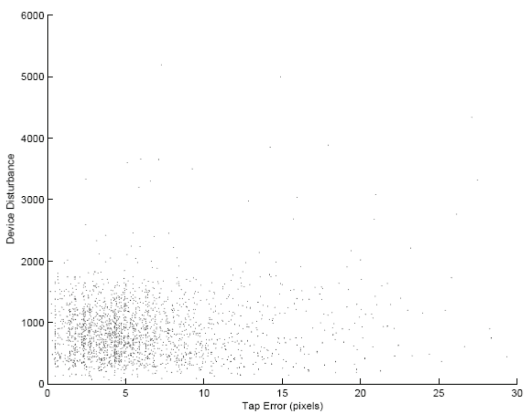


Figure 16. A scatter plot of the magnitude of the device disturbance plotted against the tap error in pixels



Other Analysis

The results presented so far have involved analysing the acceleration trace to extract information about the users' steps. Now we examine all disturbances affecting the device. If the device is moving around more, we would expect the user to be tapping less accurately. By looking at the magnitude of the of the acceleration trace in *x*, *y*, and *z* we gain an insight into the mean magnitude of disturbance that the device was going through. Figure 16 shows a scatter plot of the magnitude of the tap error plotted against the mean magnitude of the disturbance of the device for the one second previous to the tap. As the figure shows, in this instance there is no simple correlation between tap error and device acceleration.

Discussion

Using standard usability metrics, it was able to be shown that tapping accuracy was, unsurprisingly, typically greater when sitting still, rather than walking. However, the results demonstrate the extra insights into user behaviour that were made possible by taking an instrumented usability approach.

Specific experimental observations of this instrumented usability approach are:

- Users' tapping time is significantly correlated with gait phase angle. Users were approximately three times more likely to tap at the most favoured tap phase than the least favoured tap phase.

  Users' tapping position accuracy is significantly higher (lower mean error and lower variability) at these preferred phase angles. Analysis of the timing data for the different phase regions showed that users subconsciously delayed their target selection in order to tap in one particular phase area rather than any other. There is further structure in the left step-right step tap density, error biases and variability, but even when averaged over all steps, the results are significant.
- The distribution of tapping errors varies both with phase of step and between walking and sitting and across different screen positions.

  It is interesting to note that although there is no simple correlation between tap error and device acceleration, the inferred phase angle, which is based solely on the acceleration observations, does show a strong link between acceleration and tapping accuracy, emphasizing the need for appropriate models in data analysis. One potential reason for this is that the walking route chosen for the study did not require the user to make irregular adjustments to their movements. The path was quiet so that the user only infrequently had to avoid objects. This limited the disturbance of the device to lower than might be expected in a more crowded environment, or for example, in a moving vehicle. The participants grounded their tapping hand on the device while tapping, which minimised the effect of the external disturbance in this instance, so the main disturbance came from the gait cycle of walking itself.

CONCLUSION AND FUTURE WORK

This work has demonstrated that by making fine-grained observations from sensors during a usability study, that increased detail about the timing and error rates for users can be learned. Until now, linking the analysis of, for example, walking behaviour, in a realistic setting would typically have required the use of hand scoring videotapes of users' actions—a time-consuming and potentially subjective and error-prone approach which is also not open to online experimental control. Recent rapid developments in mobile device capacity and compact sensors, coupled with the use of the analytic tools from synchronization theory, have opened up a new way of investigating gait effects in interaction. The inertial sensors monitor walking patterns throughout the experiment and can potentially be used together with machine learning classification algorithms during the experiment to control for experimental stimuli and adapt the experimental situation online, providing a more stringent method of exploring mobile interaction.

The work opens new directions in both design and usability areas for future work. The specific results gained through the use of the accelerometer data for gait analysis allow the exploration of new areas to inform mobile design. For example, one question raised from this study is—does designing an interface such that users tend to tap in preferred phase ranges lead to quantitatively better performance and qualitatively more pleasant user experience? Might it be better to delay user prompts until a particular phase region, in order to sustain rhythmic interaction? (See Lantz and Murray-Smith, 2004 for a discussion of rhythmic interaction). This suggests experiments deliberately timing the presentation of prompts, or by using rhythmic vibrotactile or audio feedback in such a way that the user is pushed towards tapping in the specific phase regions. This sensor-conditional feedback can be generalised such that specific interventions can be generated in usability experiments, with a frequency which is proportional to the probability of different contexts, allowing users to 'interact in the wild' while retaining an increased level of experimental control.

The effects of bias and correlation in tapping errors can be systematically compensated for in real time, improving the tapping accuracy. This information can also be used to automatically adapt screen layout to walking speed, simplifying and spreading out the targets as the speed increases.

Further to that, there is opportunity to couple the more objective methods of measuring walking speed used in this chapter with the existing literature relating usability to the subjective use of Percentage Preferred Walking Speeds in, for example, Pirhonen et al. (2002). For experimental environments that are more difficult for a user to navigate (such as crowded streets), these techniques could potentially provide more information about user disturbances and behaviour. The online recognition of context or situations could be used to have more targeted experiments in realistic environments where a particular stimulus could be presented when the sensors recognise data compatible with a pre-specified situation.

The experiment described here specifically examines user performance when walking. However, the general approach is applicable to mobile usability studies in general as a method of gaining more information about the moment to moment actions of the user. Specifically, it allows the gaining of greater insight into user actions in an uncontrolled environment, allowing mobile usability tests to more easily take place in more realistic, less laboratory based circumstances. This work has relevance for tasks such as text entry or menu navigation in mobile settings. While this work was tap-based, similar features might be found in button-pressing, graffiti gestures, or tilt-based interaction.

ACKNOWLEDGMENT

AC & RM-S are grateful for the support of SFI grant 00/PI.1/C067 and the HEA funded *Body Space* project. RM-S, SB & BM are grateful for the support of EPSRC grant GR/R98105/01 and IRSCET BRG project SC/2003/271, *Continuous Gestural Interaction with Mobile Devices.*

REFERENCES

Barnard, L., Yi, J. S., Jacko, J. A., & Sears, A. (2005). An empirical comparison of use-in-motion evaluation scenarios for mobile computing devices. *International Journal of Human-Computer Studies, 62*, 487–520.

Brewster, S. A. (2002). Overcoming the lack of screen space on mobile computers. *Personal and Ubiquitous Computing, 6*(3), 188–205.

Crossan, A., Murray-Smith, R., Brewster, S., Kelly, J., & Musizza, B. (2005). Gait phase effects in mobile interaction. In *Proceedings of ACM SIG CHI 2005* (pp. 1312–1315). Portland, OR.

Fishkin, K. P., Jiang, B., Philipose, M., & Roy, S. (2004). I sense a disturbance in the force: Unobtrusive detection of interactions with rfid-tagged objects. In *Proceedings of UBICOMP 2004* (pp. 268–282).

Gemmell, J., Williams, L., Wood, K., Lueder, R., & Bell, G. (2004). Passive capture and ensuing issues for a personal lifetime store. In *Proceedings of ACM CARPE 2004* (pp. 48–55).

Hinckley, K., Pierce, J., Horvitz, E., & Sinclair, M. (2005). Foreground and background interaction with sensor-enhanced mobile devices. *ACM Transactions on Computer-Human Interaction, 12*(1), 31–52.

Kern, N., & Schiele, B. (2003). Multi-sensor activity context detection for wearable computing. In *Proceedings of the European Symposium on Ambient Intelligence*, Eindhoven, The Netherlands.

Kjeldskov, J. & Stage, J. (2004). New techniques for usability evaluation of mobile systems. *International Journal of Human-Computer Studies, 60*, 599–620.

Lantz, V., & Murray-Smith, R. (2004). Rhythmic interaction with a mobile device. In *Proceedings of NordiCHI '04* (pp. 97–100). Tampere, Finland: ACM.

Mizobuchi, S., Chignell, M., & Newton, D. (2005). Mobile text entry: Relationship between walking speed and text input task difficulty. In *Proceedings of MobileHCI'0* (pp. 122–128). ACM.

Morrison, A., Tennent, P., & Chalmers, M. (2006). Coordinated visualisation of video and system log data. In *Proceedings of the 4th International Conference on Coordinated and Multiple Views in Exploratory Visualization*, London

Morrison, A., Tennent, P., Williamson, J., & Chalmers, M. (2007). Using location, bearing and motion data to filter video and system logs. In *Proceedings of the 5th International Conference on Pervasive Computing* (pp. 109-126), Toronto.

Oulasvirta, A., Tamminen, S., Roto, V., & Kuorelahti, J. (2005). Interaction in 4-second bursts: the fragmented nature of attentional resources in mobile HCI. In *Proceedings of CHI 2005* (pp. 9 19–928). ACM.

Petrie, H., Furner, S., & Strothotte, T. (1998). Design lifecycles and wearable computers for users with disabilities. In *Proceedings of the First Workshop on Human-Computer Interaction with Mobile Devices.* Glasgow, UK. Glasgow University.

Pikovsky, A., Rosenblum, M., & Kurths, J. (2001). *Synchronization: A universal concept in nonlinear sciences.* Cambridge University Press.

Pirhonen, A., Brewster, S. A., & Holguin, C. (2002). Gestural and audio metaphors as a means of control for mobile devices. In *Proceedings of ACM CHI 2002* (pp. 29 1–298). Minneapolis, MN: ACM Press Addison Wesley.

Roto, V., Oulasvirta, A., Haikarainene, T., Kuorelahti, J., Lehmuskallio, H., & Nyyssönen, T. (2004). *Examining mobile phone use in the wild withQuasi-experimentation* (Tech. Rep. 2004-1). Helsinki Institute for Information Technology, Finland.

Williamson, J., Murray-Smith, R. & Hughes, S. (2007). Shoogle: Multimodal excitatory interaction on mobile devices. In *Proceedings of ACM SIG CHI Conference*. San Jose,

Yi, J. S., Choi, Y. S., Jacko, J. A., & Sears, A. (2005). Context awareness via a single device-attached accelerometer during mobile computing. In *Proceedings of MobileHCI'05* (pp. 303-306).

KEY TERMS

Accelerometer: A sensor for measuring accelerations. These accelerations can be due to both translation or rotation of the sensor. The acceleration signal can be used to measure the orientation of the sensor, but only when it is held static.

Context Aware Computing: Computer systems that can sense some aspects of the context of interaction and potentially adapt themselves accordingly.

Hilbert Transform: For an oscillatory signal s(t), the Hilbert transform H(t) gives the instantaneous phase and amplitude of s(t).

Instrumented Usability Analysis: The use of sensors during a usability study which provide observations from which the evaluator can infer details of the context of use, or specific activities or disturbances.

Oscillatory Systems: A system exhibiting periodically repeating changes of state.

Phase Angle: The phase angle gives a measure of how far an oscillating system is currently through it's oscillation, from 0 to 2PI.

Rhythmic Interaction: Interaction based upon repeated movements where both actions and the timing of the actions are important to the interpretation of the meaning.

Synchronisation: An adjustment of rhythms of two or more oscillatory systems due to their weak interaction.

Chapter LVI
Three Eye Movement Studies of Mobile Readability

Gustav Öquist
Karolinska Institutet, Sweden

ABSTRACT

Readability is important for mobile usability since text is the preferred way of dealing with information over distances in time and space. Making it easy to read on mobile devices has, however, proved difficult, primarily since the accustomed way to present texts is at odds with the limited screen space available. Simply enlarging the screen is an obvious solution, but since mobile devices need to be small to be mobile, this is just not viable. Instead, ways need to be found to present texts on small screens that facilitate the level of readability we are used to. To be able to do this, methods for evaluating novel text presentation formats on mobile devices are needed. This chapter presents findings and experiences from three readability studies where eye movement tracking has been used to learn more about how to improve readability on mobile devices.

INTRODUCTION

We live in revolutionary times. Information technology has radically changed both how information is shaped, how we work with it, and last but not least, how access to it is gained. Written language has ceased to be bound by the physical surface words which are scribed upon and have transitioned into a virtual realm. Now, texts are fundamental for the use of computers since almost all services and applications are based upon it in one way or another. With the introduction of the Internet, the greatest challenge in information access has become to find what you are looking for rather than how to gain access to it (Sahami, Mittal, Baluja, & Rowley, 2003). With the introduction of network connected mobile devices, such as mobile phones and personal digital assistants (PDAs), any electronic text can be displayed on any screen, anywhere and anytime. The mobile Internet has for several years been predicted to be the next big thing in how information will be accessed, and the predictions are well founded. Today there are over 2.5 billion users of mobile phones globally; the figure is more than twice as large as the number of Internet users and growing at a rate of 40 million per year (Wireless Intelligence, 2006). In a few industrialized countries there are now more mobile phones than citizens, and in many developing countries, a mobile phone will probably be the first computational device that most people will come to own (GSM Association, 2006). However, regardless of the fact that mobile devices are readily available and most of them are network connected, the predicted success of the mobile Internet has so far not been realized.

There are many different reasons for the slow uptake of the mobile Internet. Service providers made a mistake by promising too much too soon when claiming that the Wireless Application Protocol (WAP) was the same thing as the Internet, which it never was and never will be. Producers of mobile devices made a mistake by not providing opportunities for third party developers to create applications. Companies have not been keen to invest in mobile solutions due to lacking standards and business models. Consumers have not thought of mobile phones as computers, but rather as phones for making calls. Developers, whom actually have thought of mobile phones as computers, failed to recognize that reusing desktop interaction methods might not be the ideal solution in a mobile setting. However, all of this is changing. Service providers now offer access to the real Internet and third party developers can create applications; something that combined makes it possible for companies to build upon existing standards to extend existing, or invent new, business models. The companies can then market these services to consumers, which make them realize that their phone can be used for so much more than talk. Developers have begun to explore and utilize the novel interaction possibilities offered by mobile devices, but there is still much that remains to be done. Challenging how things are done today is the first step towards succeeding in doing it tomorrow.

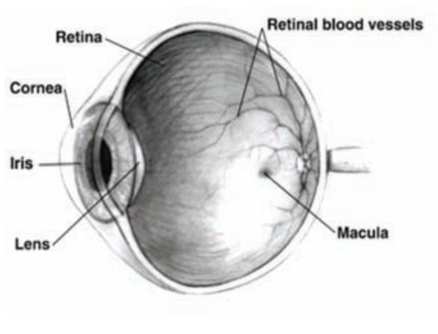
When wireless phones were enhanced with computational functionality and became Internet connected they inherited the interaction methods and office metaphors of the direct manipulation paradigm, justifiably so since these tools already had proved to be extremely useful for interaction with computers (Schneiderman, 1982). Nonetheless, the usefulness of any tool is dependent on a combination of design and use, and these tools were never intended for interaction with devices based on a design derived from mobile phones, moreover, used in a nomadic environment. It is important to see that limited input and output capabilities due to smaller keyboards and screens are not optional, they are a prerequisite since mobile devices have to be small to be mobile (Öquist, Goldstein, & Chincholl, 2004). This chapter focuses on the conflict between how text is traditionally presented, which requires a fairly large area to draw the text upon, and the limitation mobile devices put on the screen size available for this. One approach to overcome the size constraints may be to design interfaces that utilize the possibilities offered by mobile devices to dynamically work with the text and present it in a more suitable way for the user. Any such new presentation format must, however, still adhere to the principles for reading that has evolved over time. Moreover, to be able to see if the novel formats work, methods are needed to empirically evaluate them in usability studies. This chapter presents three usability studies where novel text presentation formats were evaluated on mobile devices. Apart from looking at factors such as efficiency and task load, eye tracking was used as tool to learn more about the formats. Although the results are interesting, the focus of this chapter will be on the methods used in the evaluations.

Any new format wanted to be used for text presentation must conform to how we are used to reading. Regardless of the device used for reading, or the format used for presentation, the physiological and cognitive limits for reading remain the same. A natural starting point for this chapter may, therefore, be an overview of the reading process and a clarification on what is meant by readability. This is followed by an introduction to text presentation on small screens and the merits and pitfalls of the most common approaches. Next, previous evaluations of readability on mobile devices are presented and the methods used in these are discussed. Thereafter, the eye movement studies are presented together with a comparative review of the findings. The results and the methods used to reach them are then discussed, and finally, a few concluding remarks wrap up the chapter.

THE READING PROCESS

Reading is a skill that lies deeply embedded within in the mind. Researchers have proposed several different models of how the reading process works; some are highly detailed whereas others are more generalizing. Most agree that the process can be seen as a form of pattern recognition, but most also acknowledges that exactly how process works within the brain remains to be discovered (Reichle, Rayner, & Pollatsek, 2000). Neuroimaging techniques such as Positron Emission Tomography (PET), and more recently Functional Magnetic Resonance Imaging (fMRI), have been able to show where and when processing takes place during reading (Shaywitz, Lyon, & Shaywitz, 2006). This is important as it gives a physiological understanding of the process, but it does not really answer the question about how it works. Since not much is known about the processing, a better starting point for understanding how reading is done, might be to look at the recognition. By observing how the eyes move while reading, we can tell how the recognition part works; if that is done we might also learn some about the processing as well. Next, the eye and its physiology will be closer looked at. This is followed by an overview of what is known about cognitive processing, and how eye movements in this respect are interpreted. Finally, the concept of readability and how it lends itself to measurement will be discussed.

Figure 1. The eye (courtesy of U.S. National Institute of Health)



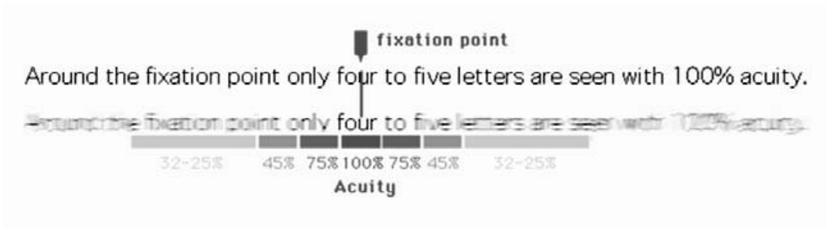
Physiological Limitations

The receptive part of the eye, called the retina, is essentially a panel full of photosensitive receptors located on the back of the eyeball (Ø ~42 mm) (Figure 1). The retina has two types of receptors, cones and rods. Cones register luminosity and colors whereas rods register light changes. Rods are much more sensitive to light, but they cannot detect colors and are also slower to respond. Most of the cones are located in a tiny area at the centre of the retina called the fovea (Ø ~0,2 mm) (located in the macula in Figure 1). The fovea is surrounded by the parafovea (Ø ~3 mm); in this region there are still many cones, but also an increasing amount of rods. Outside the parafovea there are few cones and a decreasing amount of rods, therefore vision becomes progressively less clear in the periphery of the retina (Procter & Procter, 1997).

When reading, the text must be projected on the fovea since a high concentration of cones is required for accurate recognition. Even though the retina has a 240 degree field of vision, the foveal field of vision is only 2-3 degrees wide which means that only 6-8 characters can be seen clearly in a single gaze (Robeck & Wallace, 1990). Moving centrifugally out from the fovea, the number of cones diminishes rapidly. The area immediately surrounding the fovea, the parafoveal region, further extends how much of the text that can be seen in a single fixation to around 12-14 characters (Robeck & Wallace, 1990), but beyond that the resolution is too low for recognition (Figure 2). The perceptual span is centered to the right of the fixation point, at least for readers of left-to-right languages (Just & Carpenter, 1980).

The effect of this on reading is that a very narrow focal point of vision has to be moved across the text to be able to read it. Information is processed in fixations, for example, the fixed gazes, with a

Figure 2. The perceptual span

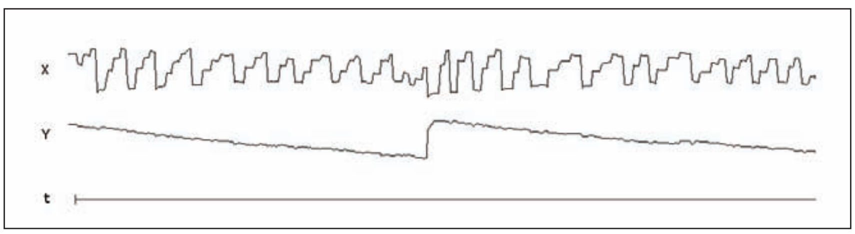


duration ranging depending on what the target is. The durations have been found to vary greatly. In some studies it has ranged between 100-500 ms (Rayner, 1998), whereas in others it has been found to vary between as much as 50-1500 ms (Just & Carpenter, 1980). To move between fixations, the eye performs very swift eye movements, called saccades, stretching up to 1-20 characters. The planning and execution of a saccade is based on the previous fixations and that which can be seen in the parafoveal region (Robeck & Wallace, 1990). Approximately every fifth saccade is directed backwards in the text for normal readers, the reason for this is that the reader has to go back and reread a word or change position within a word. When reading a text on a page with a traditional layout like this, return sweeps are used to move between the lines and page sweeps are used to move between pages (Figure 3). You can experience this for yourself; just hold a fingertip lightly on top of an eyelid and you should be able to feel how the eye moves while you read.

Cognitive Processing

The planning of saccades and the use of regressions for clarification seems to indicate that there is more to reading than meets the eye. The large differences observed in saccade lengths and fixation durations appear to reflect an ongoing process that changes depending on what is being read. What is known about the physiology of the eyes and their movements while reading seem to suggest that perception and recognition is highly dependent on cognitive (i.e., linguistic) processing. Eye movements can tell surprisingly much about cognitive processing and most models of the reading process are based on empirical data of reading. The models that have been proposed can be roughly divided into either ocular motor or processing driven models (Reichle et al., 2000). The ocular motor models mostly look at the visual properties of the text (i.e., word lengths) and the physiological limits of the eye (i.e., perceptual span and saccade lengths) in order to determine the location and duration of fixations (Reichle

Figure 3. XY-plot of reading over two pages, horizontal movements (top) and vertical (bottom) over time



et al.). Ocular motor modeling has successfully been used to predict eye movements, but the models can never (and do not claim to) explain the whole reading process since they ignore the fact that language evidently has an impact on reading. The processing models on the other hand assign linguistic processing a very central role. The general assumption of these models is that the fixation duration is directly related to the cognitive processing whereas the fixation targets are determined by a combination of linguistic, orthographic, and ocular motor factors (Reichle et al., 2000).

Just and Carpenter suggested that, "a reader can take in information at a pace that matches the internal comprehension process" (Just & Carpenter, 1980). From this starting point they developed Reader, the most widely known processing model. They began by observing actual gaze durations, the sum of all fixations on a word before moving to the next, made by college students reading scientific passages of text. Just and Carpenter found large variations in the duration of individual fixations as well as the duration of fixations on individual words. They also found that almost each content word was fixated and that fixation times were longer on words that were infrequent, thematically important, or clarifying the interpretation of previous words. The gaze durations were also found to be longer at the end of a sentence thus indicating integrative processing. From these findings they founded their model on two assumptions. The first was the immediacy hypothesis, which state that each word is immediately processed when it is fixated. The second assumption is the eye-mind hypothesis, which state that the eyes remain fixated on a word as long as it is processed (Just & Carpenter). Both assumptions have later been criticized because they do not account for context and parafoveal preview effects (Robeck & Wallace, 1990). The model presented by Just and Carpenter is very comprehensible but unfortunately it tries to explain the entire reading process, from fixation to long term-memory. Although this made the model quite complex it is still disputed as it is assumed to simplify matters too much (Reichle et al., 2000).

However, although the model might have tried to cover too much of the reading process it still has merits, simplifying a complex problem is not necessarily negative. If Just and Carpenter's processing model is combined with ocular motor modeling of the physiological limits of the eye and the visual properties of the text, we may get closer to a realistic definition. Fixation duration, that is, determination of when, is governed by cognitive processing, while saccade execution, that is, determination of where, is governed by a combination of linguistic, orthographic, and ocular motor factors. There are also other processing models available but these have had a minor impact on the work presented here. Rayner's E-Z reader is a model that is similar to the one presented by Just and Carpenter, but with a narrower scope since it does not try to account for high-level linguistic (e.g., semantic) processing. It does, however, account for preview and context effects (Rayner, 1988; Reichle et al., 2000). The model is unfortunately quite complex and the underlying assumptions are not as transparent as in the model presented by Just and Carpenter. Another processing model is the attention-shift model (Reilly, 1993). It utilizes two connectionist back-propagating neural networks, one for word recognition and one for planning saccades. From a linguistic viewpoint, the attention-shift approach seems a little too simple to be plausible; however, the use of a learning algorithm is appealing since individual differences in reading behavior are likely to be quite large.

Measuring Readability

Readability is typically referred to as the ease of "which the meaning of text can be comprehended" (Mills & Weldon, 1987). This is, of course, a very vague definition, but the assessment of readability is also affected by a multitude of factors. First, there are many differences between texts, some are very comprehensive and well written whereas others can be totally unreadable. Second, there are differences between readers; some are very experienced whereas others can barely read at all. Third, there are differences between reading situations, reading reference literature before an exam differs a lot from reading a novel while waiting for the bus. Fourth, there are differences between the presentation formats, this text might be comfortable to read on paper but is likely to be strenuous to read on a flickering screen with low resolution. To summarize: There are so many factors that affect readability that it is impossible to account for them all.

Since readability is hard to quantify, the solution is to use approximate measures instead. The readability estimations used in this chapter can be categorized according to their use as either ratings or measures. Ratings are used to determine readability of text based on quantitative predictions whereas measures are used to evaluate readability based on actual reader performance. Readability of text is usually rated by using readability formulas. Most readability formulas are quite simple and use a combination of word frequencies, word lengths, and sentence lengths as a basis for the results. Although most formulas use purely quantitative measures, they can give an indication of how hard or easy a text is likely to be to read. There are several readability formulas for English available (see Tekfi, 1987 for an overview), but for Swedish there is only one that is widely known. LIX (Läsbarhetsindex or Readability Index in Swedish) is a quantitative readability formula developed by Björnsson (1968). An estimated value of the readability of a text is calculated on basis of the percentage of long words, seven or more characters, and the average sentence length. The result is a value between approximately 1 and 100 where lower values are interpreted as easier to read. Readability has mostly been measured in terms of reading speed and comprehension (Mills & Weldon, 1987). Reading speed is often calculated as words read per minute (wpm) whereas comprehension is represented as percent of correctly answered questions. The reading speed results are mostly reliable when comparing results from different evaluations whereas the comprehension scores are unpredictable since they are highly dependant on the type of questions asked. The product of reading speed and comprehension scores are commonly used as a composite measure of reading efficiency (Jackson & McClelland, 1979; Rahman & Muter, 1999; Castelhano & Muter, 2001). The measure is used to avoid problems associated with assumed trade-offs between speed and comprehension (Wickens, 1992). However, since the comprehension scores are likely to be unreliable, both reading speed and comprehension must be reported separately as well if the results are to be comparable to other studies.

Since readability is an inherently subjective measure, subjective inventories have to be included in order to learn about the reading experience. The most widely used subjective measure is the attitude inventory. It is especially common to use when different text presentation formats are compared against each other. Attitude inventories are essentially a set of questions about experience and preference. Unfortunately, the questions often differ between evaluations, making it hard to compare the results; nonetheless they can be very illuminating for the evaluators. Another subjective measure used in evaluations, which actually is comparable, is the standardized NASA-TLX (Task Load Index) task load inventory (Hart & Staveland, 1988). The inventory is composed of six factors denoting cognitive demands that are rated by the subjects after completing a task. The inventory covers mental, physical, and temporal demand, as well as perceived effort, frustration, and performance.

What we really want to learn by measuring readability in our studies is to find the text presentation format that best supports reading. A more accurate definition of readability that relates closer to how we actually read may thus be: "the ease with which the reading process can proceed" (Öquist, Sågvall-Hein, Ygge, & Goldstein, 2004). This is where tracking and analysis of eye movements becomes interesting as this can be used as an additional measure of how the reading process has proceeded. The fact that more difficult texts or a text read in a second language resulted in a significantly larger amount of regressions was one of the first findings made after eye movements in reading were discovered a century ago (Paulson & Goodman, 2000). Looking at eye movements and observing how they differ or conform to what is expected from reading is just as useful today as it was then since it is one of the very few objective measures of readability that are available.

READABILITY ON MOBILE DEVICES

Until novel technologies such as folding screens or retinal projection prove to be viable solutions for mobile users, billions of people will read on screens with a size that is seen on mobile phones today. Since better readability results in increased usability, design guidelines for the improvement of readability based on empirical findings are important for mobile development. This section begins by looking back towards how screen reading in general has evolved with the aim of finding parallels to the future of mobile displays. Then, the presentation formats most commonly used on mobile devices today will be looked at. This is followed by a review of previous evaluations and a few remarks on evaluation methodology.

Reading on Small Screens

Readability was a problem in the early days of computing. Reading speed was found to be 20-30% slower on screens although comprehension was roughly the same (Muter, Latrémouille, Treurniet, & Beam, 1982; Kang & Muter, 1989). These findings are not too surprising given that the first screens were primitive Cathode Ray Tube (CRT) units with low resolution and mediocre refresh rates. The designs of the early experiments have also been criticized mainly because the reading situations were quite unrealistic (Dillon, 1992). The Achilles' heel of the first-generation large screens seems to have been the low resolution. Screen technology evolved rapidly and the second-generation CRT screens offered far better resolution, and also color. However, the breakthrough in readability, and usability in general, came with the introduction of the Graphical User Interface (GUI). Studies performed on computers with GUIs showed that there were in fact little or no differences between screen and paper, provided that attention was paid to such factors as screen resolution, refresh rates, anti-aliasing, text polarity, and so forth (Gould, Alfaro, Finn, Haupt, & Minuto, 1987; Osborne & Holton, 1988; Muter & Maurutto, 1991; Muter, 1996). Although reading speed and comprehension does not differ much between high-quality screens and paper, the users still seem to prefer reading on paper. This may be partially due to the fact that reading on a large screen requires the reader to view the text from a distance and in a fatiguing posture (Schneiderman, 1998). However, the screen must not necessarily be seen as a successor to paper but rather as a complement. There are many things that can be done with a text on a computer that is hard, or impossible, to do on paper. A common situation is also to browse for documents on the screen and then print the selected document on paper for reading. In this light, the readability on large screens seems quite satisfactory today. However, reading on small screens is a different story.

Most mobile devices utilize flat liquid crystal display (LCD) screens. The early LCD screens were monochrome and offered poor resolution, a bit like returning to a sized down version of an early CRT screen. However, LCD technology has evolved and today thin-film transistor (TFT) technology offers a resolution and colour depth that is comparable or better than second-generation CRT screens. The problem with readability on small screens is not so much the resolution as it is the limitation of screen space, which restricts the amount of information that can be presented at a time. This implies a higher rate of interaction by the user to view the text. Reading a longer text on a small screen can thus be frustrating and, to complicate matters further, users of mobile devices do not always have access to printing facilities. Duchnicky and Kolers (1983) performed an experiment with varying window widths and heights on a desktop computer and found that a height of 20 lines only increased reading speed by a mere 9% compared to using a height of four lines. Smaller window heights than four lines were, however, found to be significantly less efficient to use. The results also showed that a window width of two-thirds of a full page increased reading speed by 25% compared to using one-third of a full page.

These widths are much larger than what the average mobile device has to offer, but the findings seem to suggest that a limited screen width decreases reading speed. Using a higher density of characters per line was also found to improve readability; using 80 characters compared to 40, increased reading speed by 30%. This is not very surprising given that a lower density implies less information in the perceptual span at a time, and therefore, also a lowered efficiency. Dillon, Richardson, and McKnight (1990) investigated how reading was affected by using window heights of 60 and 20 lines. The subjects who read using the smaller window height were found to perform significantly more jumps and also altered the direction of reading much more often, but the results showed that neither reading speed nor comprehension differed. This is not in line what Duchnicky and Kolers found, one reason for the difference may be that Dillon used much longer texts (~3000 words compared to ~350 words). Nonetheless, the results from both evaluations seem to suggest that a small screen space does imply a higher rate of interaction.

Text Presentation Formats

Since reading on computer screens is a recent innovation, it is natural that the two most commonly used formats to present text, scrolling and paging, are adapted from how text on papyrus and paper are traditionally presented. Computers do, however, offer new ways to present text on screens by displaying it dynamically over time (Bruijn & Spence, 2000; Juola, Tiritoglu, & Pleunis, 1995; Mills & Weldon, 1987; Rahman & Muter, 1999). The two most well know formats for doing this are leading and Rapid Serial Visual Presentation (RSVP). Each of the four formats, scrolling, paging, leading, and RSVP, will now be presented in turn with focus on how the may be used on mobile devices.

Scrolling

Scrolling presents the text in the traditional format on a display area that may be larger than the screen. Scroll bars are usually used to indicate how much of the text that is displayed on the screen as well as current horizontal and vertical position. Using both horizontal and vertical scrolling for text presentation is neither efficient nor appreciated by users (Muter, 1996), the width of the lines are, therefore, often shortened to fit the screen so that only vertical scrolling is necessary. A joystick or a set of arrow keys can be used to move to a different position. In interfaces where a mouse or stylus is used, the scrollbars may be used for navigation as well. On mobile devices, the text is usually vertically scrolled line by line. The number of lines is decided by the amount of text that can fit horizontally. The least number of interactions (e.g., button presses, etc.) needed to read a text using scrolling is equal to the number of lines minus the number of lines initially displayed on the screen (Figure 4).

Paging

Paging presents the text in the traditional format, but divides it into pages that fit the screen area. A joystick or a set of arrow keys is used to move between pages. The current page number and the total number of pages are usually displayed to inform the user of the position in the text (Muter, 1996). The number of pages needed to present a text is a function of the line length, decided by how much that can fit horizontally, and the number of lines per page, decided by how much that can fit vertically. If either the line length or the

Figure 4. Scrolling implemented on a mobile phone

number lines per page is small, as it usually is on mobile devices, even slight changes to either can dramatically change the number of pages. The least number of interactions needed to read a text using paging is equal to the number of pages minus the first page displayed. Compared to scrolling, the number of interactions is thus reduced by a factor of how many lines that fit per page on the screen (Figure 5).

Leading

Leading dynamically scrolls the text horizontally on one line across the screen. The text moves continuously across the screen at a certain speed that may be selected by the user. Moving the text pixel for pixel has been found to be more efficient than moving it character for character (Kang & Muter, 1989). A joystick or a set of arrow keys are used to start and stop the presentation, go forward and backward in the text, as well as increasing or decreasing the speed. A progress bar or completion meter may be used to indicate location in the text. The time required to read a text is decided by the speed of the text presentation. The size of the text display area does not affect the presentation speed, but text presented on a shorter line may be perceived as going faster. The least number of interactions needed to read a text using leading is just one (or even none if the presentation starts automatically) but speed changes as well as interactions to go back and forward realistically adds to this number (Figure 6).

Figure 5. Paging implemented on a mobile phone

Figure 6. Leading implemented on a mobile phone

Rapid Serial Visual Presentation

Rapid Serial Visual Presentation dynamically presents the text in chunks of one or a few words at a time, at a fixed location on the screen (Forster, 1970; Juola, Ward, & McNamara, 1982; Potter, 1984). The chunks are successively displayed at a pace that may be selected by the user (Muter, 1996). Adapting the exposure time of each chunk to its length or frequency in language has been found to improve readability (Castelhano & Muter, 2001; Öquist & Goldstein, 2003). A joystick or a set of arrow keys are used to start and stop the presentation, go forward and backward in the text, as well as increasing or decreasing the speed. A progress bar or completion meter may be used to indicate location in the text and previous evaluations have shown that this is beneficial for the RSVP format (Rahman & Muter, 1999). The time required to read a text is decided by the speed of the text presentation, which usually is measured in words per minute. The size of the chunks affect the speed of the presentation since smaller chunks requires a higher presentation pace. The time to read a text is, however, constant at a certain speed regardless of chunk size. In similar to leading, the format requires a minimum of user interaction (Figure 7).

Figure 7. RSVP implemented on a mobile phone

Today most mobile phones use a joystick in combination with a set of soft keys to control the interface. The joystick is typically used to control the text presentation whereas the soft buttons are used to control the interface, for example, switch between menus or alter settings (Figure 8). Utilizing accelerometers to control the presentation by tilting the device may prove fruitful in the future as it offers a more direct mode of interactivity (Öquist, 2004).

Previous Evaluations

Most evaluations of the text presentation formats targeted for small screens have not actually been performed on mobile devices. The majority of them have, moreover, been directed towards exploring new possibilities to present text, usually by evaluating novel variations of the RSVP format. To make matters a bit more complicated, the implementations of the text presentation formats and the experimental designs vary considerably between experiments. It is, therefore, hard to compare a finding for one format reported in one evaluation to those achieved in another. Nonetheless, an overview of previous experiments may at least shed some light on what has been learned so far about the text presentation formats in terms of readability.

Juola et al. (1982) presented shorter paragraphs of text on a CRT screen, either in the page format or in the RSVP format with text chunks of 5, 10, or 15 characters. Each text chunk was exposed for 200-300 ms, which is equal to a reading speed of approximately 300 wpm. The results showed no significant differences in comprehension between the reading conditions.

Masson (1983) evaluated how the insertion of blank windows at sentence boundaries affected the RSVP format. Masson experimented with durations of 500 and 1000 ms and found that performance increased with blank windows regardless of duration.

Cocklin, Ward, Chen, and Juola (1984) compared RSVP with the text divided into either idea units or ad-hoc chunks. The idea unit segmentation was performed by hand and was based on clause and phrase boundaries as well as linguistic features. Each chunk averaged 13 characters and the reading speed was approximately 300 wpm. The results showed that the use of idea units increased comprehension a little but not significantly.

Muter, Kruk, Buttigieg, and Kang (1988) performed experiments with self-paced RSVP and RSVP that permitted regressions. The results showed that larger regressions yielded slower reading and regressions back to the beginning of the sentence were found to be more frequent than regressions two words back. Overall, the results indicated that permitting reader control was feasible but permitting regressions resulted in lower performance.

Figure 8. Button assignments for RSVP on a Sony Ericsson T610 mobile phone

Kang and Muter (1989) compared RSVP to word-by-word, letter-by-letter, and pixel-by-pixel leading. Except for word-by-word, comprehension was as high for leading as it was for RSVP. The comprehension scores for pixel-by-pixel leading were also found to match RSVP at reading speeds ranging from 100 to 300 wpm. The subjects in the evaluation were also found to express a significantly higher preference for pixel-by-pixel leading.

Fine and Peli (1995) evaluated how visually impaired and elderly subjects read using RSVP and scrolled text. They found that the visually impaired read at a similar speed using both formats whereas the elderly read faster using RSVP.

Juola et al. (1995) compared leading to RSVP on an eight-character horizontal display. The results showed that sentences were read more accurately in the RSVP format than in the leading format.

Rahman and Muter (1999) benchmarked word-for-word RSVP and sentence-by-sentence presentation, with or without a completion meter, to traditional text presentation in the page format. No significant differences were found for comprehension and reading speed but the subjects liked the inclusion of a completion meter.

Sicheritz (2001) compared reading using RSVP with three different text presentation window widths (11, 17, and 25 characters) on a PDA to reading in a paper book. The results showed that neither reading speed nor comprehension differed between the conditions. The NASA-TLX task load inventory did, however, reveal significantly higher task load ratings for the RSVP conditions for all factors but Physical demand. A 25-character window width was found to be more efficient, but the difference was not significant.

Castelhano and Muter (2001) evaluated the effects of using RSVP with or without punctuation pauses, variable word durations, and a completion meter. They compared a few RSVP formats to traditional text presentation and sentence-by-sentence presentation. The results showed that pauses and variations made the RSVP format significantly more accepted. However, the sentence-by-sentence and traditional page format remained more popular although RSVP was just as effective.

Laarni (2002) compared reading using scrolling, paging, leading, and RSVP on several different screen sizes that were emulated on a desktop. The results showed that scrolling and RSVP were the most suitable formats to use on mobile phones. RSVP was, moreover, found to be the fastest format when reading on a screen with the size of a mobile phone.

Öquist and Goldstein (2003) evaluated three different versions of RSVP (one displaying each chunk for equal time and two variants adapting the presentation speed to the content), on a PDA against scrolling (using MS Internet Explorer) for short texts and paging (using MS Reader) for long texts. The results showed that RSVP was significantly faster compared to scrolling but more demanding to use compared to paging. The results, moreover, showed that adaptation significantly decreased NASA-TLX task load ratings for most factors.

Evaluation Methodology

All the findings reported in the reviewed evaluations are valid in the sense that they are reached scientifically. For each experiment, hypotheses are tested statistically using methods appropriate for their respective experimental design. All of the experiments are more or less repeatable; some of the software used in the experiments may be hard to come by. The main problem is, however, that none of the experimental designs are comparable to each other. This makes it very difficult to compare the results. The most commonly evaluated format in the experiments is RSVP, but not a single evaluation have implemented the format it in the same way. All use words per minute as a measure of reading speed, but it remains unclear if it is defined equally in terms of what the time include. Only a few of the comprehension tests and subjective inventories used are comparable. This does not mean that there is nothing to learn from the previous experiments, but it does mean that one has to be careful when drawing conclusions from a comparison of results. The current state of affairs is understandable given that different researchers have performed most evaluations, but it is nonetheless, regrettable. It clearly illustrates the need for standardized evaluation measures and guidelines for evaluation of readability, whether it is performed on a mobile device or not.

Evaluating readability is difficult. To start with, a representative number of subjects that read texts using the formats that you want to evaluate are needed. How large the number must be is decided by the experimental design. Running readability studies is typically time consuming, and designs that limit the number of subjects are usually desired. Moreover, there are a few things to keep in mind. Since the subjects cannot read the same text twice in the same way, one text for each text presentation format is needed. Even if the difficulty of the texts is rated and found to be similar, they are not going to be equal. A text read in one format may for some reason fit better for that format, or it may just be that the questions on the comprehension test was easier for that text. Texts and presentation formats thus has to balanced. It may come as a surprise, but the single largest source of error in a readability evaluation can be the subjects. The difference between how persons read is generally much larger than the differences between text presentation formats or the texts themselves. To ensure that it really is the text presentation formats that are being evaluated, each subject has to read a different text using each presentation format. Finally, subjects actually do get tired when reading even if it sounds like an easy task. The texts may not be that difficult, but the presentation formats that are being evaluated may be cumbersome to use. In order to get reliable results, it is important to balance the order in which subjects read using the presentation formats.

In the last of the experiments reviewed (Öquist & Goldstein, 2003), a within-subject repeated-measurement experimental design that meets all of the aforementioned requirements was used. The benefit of using such a design in readability studies is that it limits the effects of variance caused by the subjects reading performance on the results for the text presentation formats. Each subject reads a text using each of the formats, what is then looked at is not how well the reader performed but rather how well the formats performed for that reader. When several measurements are taken on the same experimental unit, in this case the different presentation formats, the measurements tend to be correlated with each other. The correlations between formats can then be taken into account using a multivariate analysis of variance (MANOVA). The repeated-measurement, general linear model (GLM) was used to test for significances. The significance (alpha) level was set to 5%. Since several hypotheses are tested simultaneously, the level of multiple comparisons has been Bonferroni adjusted (Bonferroni, 1935) (e.g., the standard of proof needed is heightened by dividing the alpha level by the number of factors). To limit the number of subjects, a graeco-latin-square design was employed. A latin-square is a table with $n*n$ cells where every element occurs exactly once in each row and column. All solutions to a Soduko are for example, latin-squares. A graeco-latin-square (also called Euler square, after the inventor) is a latin-square of two sets of n elements, S and T, ordered in a $n*n$ table so that each cell contain an ordered set $<s, t>$ and no row or column contains more than one s or one t (Box, Hunter, & Hunter, 2005). If the formats are let to be S = {A, B, C, D} and the texts T = {a, b, c, d}, a graeco-latin square (Table 1) can be created.

Table 1. Graeco-latin-square of the fourth order for presentation formats and texts

Format A / Text a	Format B / Text b	Format C / Text c	Format D / Text d
Format B / Text b	Format C / Text c	Format D / Text d	Format A / Text a
Format C / Text c	Format D / Text d	Format A / Text a	Format B / Text b
Format D / Text d	Format A / Text a	Format B / Text b	Format C / Text c

Since presentation order is also a factor to take into account, the graeco-latin-square is randomized by transposition to create three additional fourth order squares. The result is a table of 16 rows. A subject can then be assigned to each row; the experimental conditions are then the ordered cells in that row. Using a latin-square design made it possible to run a reliable experiment with four text presentation formats by employing 16 subjects. This may sound like a small figure, and in fact it is, but the statistical model used for the experiment is intended just for such situations where it is impractical or expensive to run a large number of tests. Since this experimental design seemed to work well, it has been used in all subsequent studies. However, as it will soon be seen, in the second eye movement study, tampering with the latin-square design can easily result in flawed results.

EYE MOVEMENT STUDIES

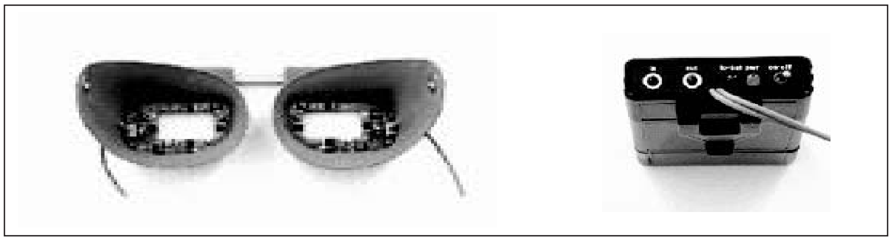
In order to learn more about how small screens affect readability, it was decided that a tool be developed that would enable eye movements to be monitored while reading on mobile devices. Using eye movements as a measure of readability connected well with the revised definition of readability, for example, "the ease with which the reading process can proceed." The main assumption was that eye movements conflicting with how we usually read could be seen as an indication of increased task load and decreased readability. Eye movement tracking has since been rewardingly used in three readability studies, two performed on a PDA and one on a mobile phone. In this section, each of the studies will be presented in turn. Focus will be on the rationale behind the evaluations, the tools, and methods involved in performing them, and of course, the key findings stemming from them.

Study One: Reading on a PDA

The aim with the first study was to observe how traditional and dynamic text presentation affected readability in terms of comprehension score, reading speed, task load rating, and eye movements (Öquist, Sågvall-Hein, et al., 2004). What was wanted is that the conditions that fared best in the previous evaluation be compared; for example, traditional text presentation in the page format and dynamic text presentation in the RSVP format using adaptation (Öquist & Goldstein, 2003). The IOTA XY-1000 eye tracking system was used for eye movement detection and integrated with the Compaq iPAQ used for evaluation. The eye tracking system consists of a pair of goggles in which infrared (IR) diodes emit light onto the eyes (Ober, 1994). The IR reflections on the eyes are sensed by eight sensors, four for each eye, which may be sampled at a frequency of up to 1 kHz. The processing unit is connected to a PC running the Orbit eye trace program, which converts the eye movements into horizontal and vertical coordinates and records them. The benefit of the system is that it is comfortable to wear, not invasive, and can record eye movements with a frequency of up to 1000 Hz. The downside is that the recordings are affected by head movement (Figure 9).

The result of a recording is a set of horizontal and vertical coordinates for the position of each eye over time. Before any recording can be made, the system has to be calibrated so that the recorded coordinates really correspond to the coordinates on the screen. To do this, a nine-point calibration pattern was displayed on the PDA and the user was asked to look at each point in turn. Only when the coordinates recorded by the eye tracker agreed with the coordinates actually looked upon could the recording start. Next, the system had to be aligned so that a known horizontal and vertical distance on the screen was available in the recording; animating a four point cross on the mobile phone with a known distance between each point did this. Given that the distance between these points is known, it is possible to calculate the position of each eye on the screen for the duration of the recording. In order to be able to calibrate and align the system with the text presentation on the PDA, a program was developed that automatically sets up an eye movement recording session, maintains synchronization with the mobile device,

Figure 9. The IOTA XY-1000 system with goggles (left) and processing unit (right)



and enables monitoring of the recording throughout the session. For text presentation, Microsoft Reader was used for the page condition whereas Bailando (Öquist & Goldstein, 2003) was used for the RSVP condition (Figure 10).

The experiment took place in a dedicated eye movement laboratory. All subjects were instructed to read at a pace that was comfortable to them and they were allowed the presentation speed at any time. While reading, the subject was seated in a comfortable chair with the head held in a fixed position by an adjustable kin support. Although this is not a very natural reading position, realism had to be sacrificed for reliable experimental data. The experimenter was seated near the subject and monitored the recordings (Figure 11). To limit the amount of data generated by the eye tracker, the sampling rate was set to 100 Hz. This is considerably lower than the system can handle, but given that reading a text takes a while, it would just be to cumbersome to deal with the data if a higher setting was used.

A balanced within-subject repeated-measurement experimental design was employed for the experiment. Two conditions were formed where each subject read one text, using either presentation format. The conditions were balanced against presentation order and texts, thus generating four combinations, which were each repeated four times, yielding 16 experimental sessions. One subject was assigned to each of the 16 sessions at random**.** No difference in reading speed, comprehension, task load, and eye movements were set as a null hypothesis. The hypotheses were tested in the SPSS V11.5 software using the repeated-measurement General Linear Model. The significance level was set to 5% and the level of multiple comparisons was Bonferroni adjusted. Sixteen

Figure 10. Interface of Microsoft Reader (left) and Bailando (right)

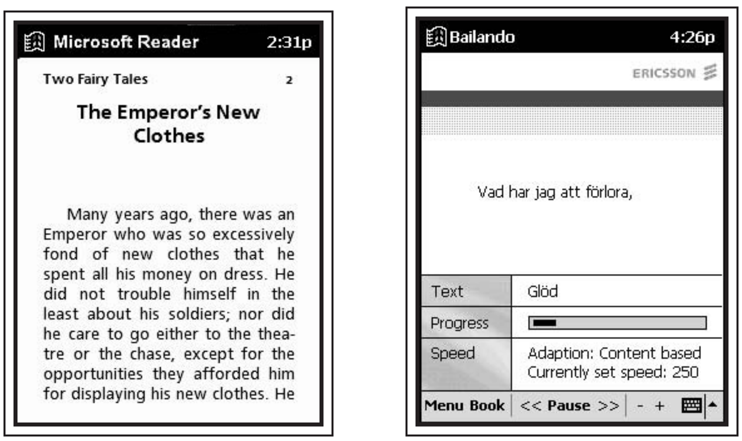


Figure 11. Setup of the first study with subject (left) and experimenter (right)

subjects (eight males and eight females; mean age: 28) participated in the experiment. Two Swedish fiction texts of similar length (~2500 words) and difficulty (LIX ~30) were chosen to be included in the experiment. Comprehension was measured for each text by ten multiple-choice questions with three alternatives, the NASA-TLX inventory was used to measure task load.

The eye movement recordings were analyzed using the JR saccade detection program (Ygge, Bolzani, & Tian, 1999). The program was used to single out movements in the recordings; eye movements were defined as continual changes in the recording with durations lasting more than 10 ms independently detected in each of the four channels (e.g., horizontal and vertical movements for both left and right eye). Using this threshold, anything else than the detected movements can be assumed to be a fixation. The movements were categorized according to their function when reading based on duration, velocity, amplitude, and co-occurrence as either: saccades and regressions (≤4 deg. without vertical movement), forward and backward sweeps (>4 deg. without vertical movement or ≤4 deg. with vertical movement), stray sweeps (>4 deg. with vertical movement), and eye blinks (peak values caused by opening and closing the eyelids). The number of movements for each category was then normalized in respect to the length of the recordings into type of movements per minute.

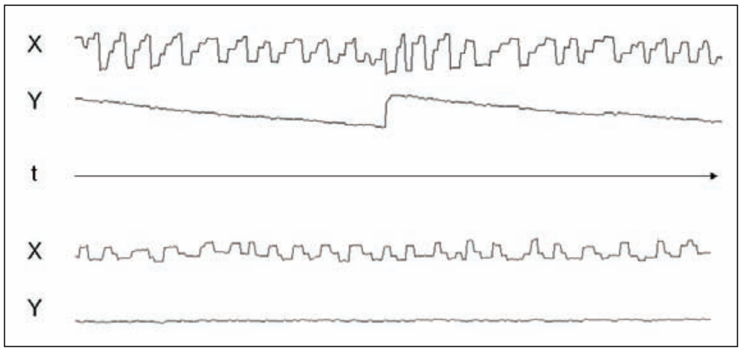
The statistical analysis indicated no significant differences in reading speed or comprehension (Table 2). However, the null hypothesis regarding no difference in task load between the conditions was rejected as there was a significant difference ($F[1,15]=25.4$, $p<0.001$). Pair-wise comparisons revealed that the use of RSVP format resulted in significantly higher ($p<0.001$) Temporal demand compared to using the page format.

The most striking differences were found in the eye movement recordings (Figure 12). RSVP was found to significantly increase the number of regressions ($p<0.001$), although it also decreased the number of saccades significantly ($p<0.006$). These findings were interesting since the advantage of the RSVP format originally was presumed to be the elimination of eye movements, which would lead to a possible reduction in cognitive load (Potter, 1984). The results show that the RSVP format does not eliminate eye movements, although it does reduce them. The reduction does, however, not seem to reduce cognitive load, it rather seems to increase cognitive load. The reason for this may

Table 2. Results for reading speed and comprehension in the first study

Condition	Reading Speed (WPM)		Comprehension (% correct)	
	Avg.	Std. Dev.	Avg.	Std. Dev.
Page format	216.9	78.7	78.1	14.7
RSVP format	191.9	45.1	74.4	20.0

Figure 12. Time-plot of a ~30 s. excerpt of eye movements for paging (top) and RSVP (bottom)



be the increase in regressions, which can be seen as an indication of when the reading process has not proceeded with ease.

These empirical findings contradict the theoretical basis of RSVP, which means that the format may have to be reconsidered. From these findings, it was suggested that a dynamic text presentation format like RSVP maybe should not try to reduce eye movements, but rather try to stimulate an eye movement pattern similar to when reading in the page format.

Study Two: Verifying the Results

The primary aim with the second study was to verify that the RSVP format really could not eliminate eye movements (Danvall, 2004). The implementation used displayed as many words as could be fitted into a chunk of 25 characters. The decision to use this chunk size was based on the findings in Sicheritz's (2000) experiment, as it showed that this was more efficient than using smaller chunks. However, the most common implementation of RSVP displays one word at a time centered on the screen. For this experiment a one word RSVP implementation that could be compared to the previous implementation was thus developed. As a mean to validate the results from the previous evaluation, it was also chosen for paging to be included in this experiment. Since the latin-square experimental design allows for four different formats, there was also an opportunity to try out a new idea. In the previous readability studies performed using Bailando on the PDA (Öquist & Goldstein, 2003; Öquist, Sågvall-Hein, et al., 2004) some of the subjects had suggested that they would like to use a RSVP format displaying more lines than one. Therefore, buffered RSVP, in which 3 chunks of 25 characters are displayed on the screen stacked upon each other with the most recent line at the bottom, was implemented. The new formats were incorporated into the Bailando prototype; eye movements were recorded and analyzed using the same hardware and software as in the previous study (Figure 13).

A balanced within-subject repeated-measurement experimental design was employed for the experiment. Four conditions were formed where each subject read one text using either presentation format. The experiment was performed as a thesis project (Danvall, 2004) and to reduce the number of subjects required, the author of this thesis decided to let all subjects read using the page format first. Only the RSVP formats were balanced against presentation order and text according to a latin-square design (sic!). This gave 12 experimental sessions to which a subject was randomly assigned. No difference in reading speed, comprehension, task load, and eye movements were set as a null hypothesis. The hypotheses were tested in the SPSS V13.0 software using the repeated-measurement general linear model. The significance level was set to 5% and the level of multiple comparisons was Bonferroni adjusted. Twelve subjects (six males and six females; mean age: 26) participated in the experiment.

Figure 13. Text presentation formats evaluated in the second study: paging (left-most), buffered RSVP (left), chunked RSVP (right), and word RSVP (right-most)

Four Swedish fiction texts by Astrid Lindgren of similar length (~1000 words) and difficulty (LIX ~30) were chosen to be included in the experiment. Comprehension was measured for each text by five multiple-choice questions with three alternatives, the NASA-TLX inventory was used to measure task load (Hart & Staveland, 1988). The eye movement recordings were analyzed using the same software and metrics as in the previous study (Figure 14).

The statistical analysis showed that the null hypothesis for reading speed could be rejected since there were significant differences (F [3,33]=7,787, p<0,001). Pair-wise comparisons showed that the page format was read significantly faster compared to chunked RSVP (p<0.033) and word RSVP (p<0.010). There were no significant differences in comprehension between any of the formats (Table 5).

The null hypothesis for task load could be rejected since there were significant differences (F[3,33]=3,317, p<0,032). Pair-wise comparisons revealed significances for all factors. Mental demand was significantly lower for the page format compared to all RSVP formats; buffered (p<0,010), chunked (p<0,012), and word (p<0,011). Physical demand was significantly lower for the page format compared to the buffered RSVP format (p<0,011). Temporal demand was significantly lower for the page format compared to both buffered RSVP (p<0.013) and chunked RSVP (p<0.002); word RSVP was, moreover, rated significantly lower than buffered RSVP (p<0.002). Performance was rated significantly higher for the page format compared to the buffered RSVP format (p<0.008). Effort was rated significantly lower for the page format compared to buffered RSVP (p<0.001), chunked RSVP (p<0.002), and word RSVP (p<0.027). Frustration was rated significantly

Figure 14. Setup of the second study with subject (left) and experimenter (right)

Table 3. Results for reading speed and comprehension from the second study (Danvall, 2004)

Condition	Reading speed (WPM)		Comprehension (% correct)	
	Avg.	Std. Dev.	Avg.	Std. Dev.
Page format	239,9	48,9	93,3	9,8
Buffered RSVP	210,2	46,9	96,7	7,8
Chunked RSVP	201,8	44,0	91,7	13,4
Word RSVP	177,0	41,1	100,0	0,0

lower for the page format compared to buffered RSVP ($p<0.001$), chunked RSVP ($p<0.004$), and word RSVP ($p<0.017$) (Figure 15).

The results for reading speed and task load were not inline with previous findings (Öquist & Goldstein, 2003; Öquist, Sågvall-Hein, et al., 2004). The page format was now significantly faster and significantly less demanding for most factors compared to the chunked RSVP format although the implementations were exactly the same. This is probably a result of the flawed experimental design. If all formats had been balanced against presentation order according to the latin-square design, something that only had required four more subjects, the findings might have been different. The comparisons between the RSVP implementations are more reliable. The buffered RSVP format that subjects had wished for in earlier experiments was probably not so good since the task load ratings were so high, interestingly enough it was quite fast, however. The chunked RSVP format was rated more demanding to use than word RSVP, a quite interesting finding given that the RSVP format was set out to be disapproved. Now, could RSVP also eliminate eye movements?

The eye movement analysis showed that word RSVP could almost eliminate eye movements (Figure 16). The findings for the page format and the chunked RSVP format were inline with the results from study two (Öquist, Sågvall-Hein, et al., 2004). The buffered RSVP format gave more vertical variation, but was otherwise similar to the chunked RSVP format. The word RSVP format significantly reduced the number of saccades and regressions compared to all the other formats ($p<0.001$) (Table 4).

Figure 15. Box plot of NASA-TLX task load ratings (Danvall, 2004)

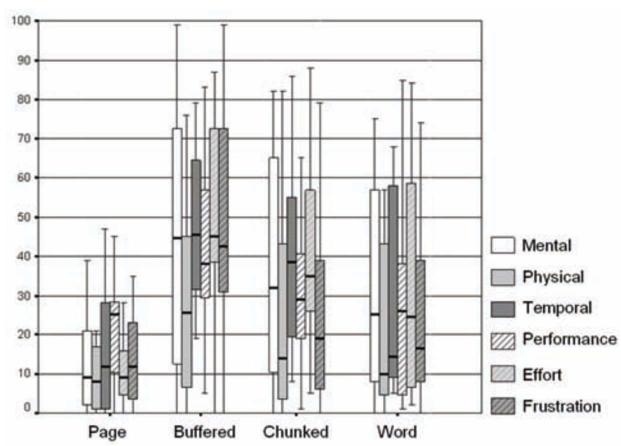


Figure 16. Plot of ~30 s. of eye movements for the same subject superimposed over the presentation formats

Table 4. Saccades and regressions per minute for left and right eye (Danvall, 2004)

Condition	Saccades per minute		Regressions per minute	
	Right eye	Left eye	Right eye	Left eye
Page format	73,6	74,2	19,3	15,2
Buffered RSVP	40,5	40,8	36,3	33,7
Chunked RSVP	43,7	44,1	40,7	36,5
Word RSVP	0,6	0,9	0,9	0,7

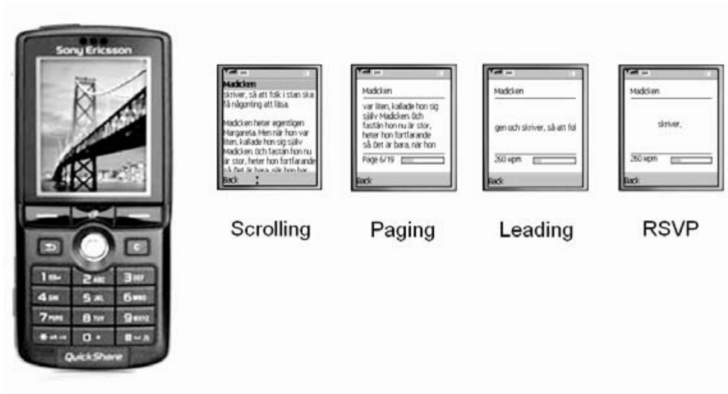
The results from this and the previous studies, show that the page format in Microsoft Reader used on a PDA is hard to beat. The RSVP format has been, with a few exceptions, equally efficient in terms of reading speed and comprehension. The problem seems to be the extra task load induced by the format. Using chunked RSVP results in a saccade/regression ratio close to 1:1 as opposed to 5:1 on paper, this may be a partial explanation, as this hardly is a natural way to read. Using word RSVP more or less eliminates the need for eye movements, but this does not seem to increase reading speed or reduce task load. The format probably really needs to be reconsidered. Maybe a dynamic text presentation format like RSVP should try to stimulate an eye movement pattern more similar to reading on paper.

However, there was also one more issue that had to be addressed before this could be tried out. All evaluations so far had been performed on a PDA with a screen much larger screen than those typically used on mobile devices. How would the RSVP format work out compared to the page format on a much smaller screen?

Study Three: Reading on a Mobile Phone

The aim with the third study was to compare traditional text presentation to dynamic text presentation on a mobile phone (Lundin, 2006). For traditional text presentation, scrolling and paging was chosen to be included as these are the formats most commonly used on mobile devices. For dynamic text presentation, leading and RSVP was included, as these are the most commonly known dynamic formats. Since word RSVP is the most common implementation, that version of the format was used. To be able to evaluate the readability of using the four text presentation formats on a mobile phone and measure eye movements, a Java 2 Micro Edition (J2ME) application was developed and integrated with the eye movement tracker used earlier. The intention with the application was to keep as many aspects of the text presentation equal as possible except for the format used. The formats were, moreover, supposed to be generic in the sense that they should be representative of how

Figure 17. The Sony Ericsson K750i mobile phone used in study three (left) and the presentation formats in the same scale (right), see Figures 4-7 for close-ups



they usually are implemented. A Sony Ericsson K750i mobile phone was chosen for the experiment since it supported J2ME and had a screen with a size typical for new mobile phones, 176 x 220 pixels (28 x 35 mm), 0.158 mm dot pitch, 18-bit (262,144 colors) TFT LCD (Figure 17).

For scrolling, the native text presentation interface offered by the phone was used (20 characters x 8 lines); the joystick or the keypad could be used to scroll up or down. For the other presentation formats, custom canvas interfaces were developed using the same font (Times New Roman 10 pixels) and screen settings (black on white). Each of the custom formats displayed the text title, a progress bar, and page numbers or speed settings. The page format displayed five lines at a time and the joystick or the numeric keypad was used to flip pages. The scrolling format displayed the text at one line in the middle of the screen and moved the text pixel for pixel. The RSVP format presented one word at a time centered on the screen; the exposure time was calculated using the content adaptive RSVP algorithm (Öquist & Goldstein, 2003). For leading and RSVP, the joystick or the numeric keypad was used to control presentation speed (up/down), going backward and forward in the text (left/right), or starting and stopping (joystick press). The initial presentation speed for the dynamic formats was always set to 250 wpm. An updated XY-1000 system offering better resolution was used in this study. Calibration was manual using a nine-point pattern on the phone whereas the alignment process was automated (Lundin, 2006).

A balanced within-subject repeated-measurement experimental design was employed for the experiment. Four conditions were formed where each subject read one text using either presentation format. The experiment was performed as a thesis project (Lundin, 2006), but wise from experience, reducing the number of subjects was not tried this time. All presentation formats were balanced against presentation order and text according to a latin-square design. This gave sixteen experimental sessions to which one subject was randomly assigned. No difference in reading speed, comprehension, task load, and eye movements were set as a null hypothesis. The hypotheses were tested in the SPSS V14.0 software using the repeated-measurement General Linear Model. The significance level was set to 5% and the level of multiple comparisons was Bonferroni adjusted. Sixteen subjects (eight males and eight females; mean age: 26) participated in the experiment. The same Swedish texts used in the previous study were used in this one but the questions in the comprehension inventory were made more difficult. Comprehension was measured for each text by five multiple-choice questions with three alternatives, the NASA-TLX inventory was used to measure task load (Hart & Staveland, 1988). The eye movement recordings were analyzed us-

Figure 18. Experimental setup, subject (left), and experimenter (right) (Lundin, 2006)

ing the same software as in previous studies, but the analysis metrics had to be adjusted to match the updated eye tracker and the new experimental setup (Figure 18).

The statistical analysis showed that the null hypothesis for reading speed was rejected since there was a significant main effect (F [3,45] = 28.35, p<0.001). Pair-wise comparisons showed that RSVP reduced reading speed significantly compared to all other formats (p<0.002) and that the page format increased reading speed significantly compared to the scrolling format (p<0.002). The null hypothesis regarding no difference in comprehension was kept (Table 5). The null hypothesis for task load between the conditions was rejected (F [3,45] = 4.26, p<0.010). Pair-wise comparisons showed that mental demand was rated significantly higher for the leading format compared to the paging format (p<0.009). Physical demand was significantly higher for the scrolling (p<0.003) and leading formats (p<0.005) compared to RSVP. Temporal demand was significantly higher for the leading and RSVP formats compared to the scrolling (p<0.001, p<0.013) and paging formats (p<0.001, p<0.011). Finally, effort was found to be significantly higher for the leading format compared to the scrolling (p<0.049) and paging (p<0.003) formats.

Unfortunately, four of the eye movement recordings were too distorted to be usable for analysis. To make matters worse, the distorted recordings were from four different subjects. The cause of the distortions was probably that the manual calibration process was too imprecise; other recordings for the same subjects were fine. Running additional subjects would have been an acceptable solution, but as both equipment and interfaces had been updated after the experiment, it was decided that upcoming experiments would be focused on instead. Since the latin-square design does not allow missing cases in the statistical analysis, only descriptive statistics can be offered for the eye movements analyzed in this study. The recordings that did work, however, presented some interesting data (Table 6).

Scrolling resulted in more vertical variations and more regressions than paging, but yielded for most aspects, fairly typical reading movements.

Table 5. Results for reading speed and comprehension from the third study (Lundin, 2006)

Condition	Reading speed (WPM)		Comprehension (% correct)	
	Avg.	Std. Dev.	Avg.	Std. Dev.
Paging	178,1	60,1	92,5	12,4
Scrolling	217,7	70,7	87,5	14,4
Leading	195,2	56,4	88,8	16,3
RSVP	135,4	44,3	92,5	12,4

Table 6. Eye movements per minute (Std. dev.) for the four text presentation formats

	Saccades	Regressions	Eye blinks	Distortions	N
Paging	74,5 (17,9)	55,9 (20,6)	19,8 (11,8)	11,83 (2,2)	12
Scrolling	76,0 (27,2)	48,7 (16,7)	28,6 (15,1)	11 (3,5)	12
Leading	93,4 (23,9)	88,46 (53,9)	14,5 (12,9)	9,62 (3,0)	12
RSVP	20,8 (11,8)	22,96 (10,8)	3,9 (2,8)	7,89 (2,5)	12

Figure 19. Scrolling eye movements

Figure 20. Paging eye movements

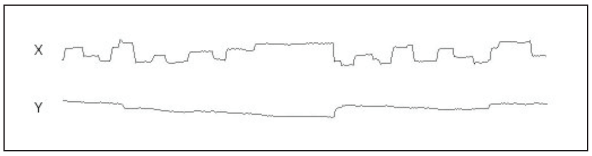


Figure 21. RSVP eye movements

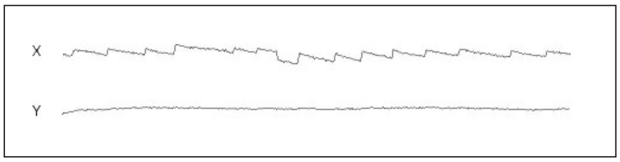
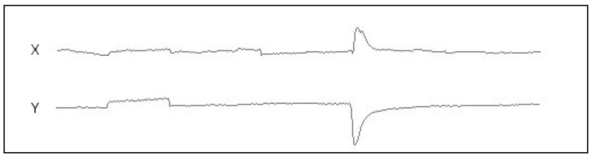


Figure 22. RSVP eye movements



The paging format resulted in very typical reading eye movements. The number of eye blinks was, however, much higher than the other formats. The leading format yielded much more eye movements than any of the other formats; it seems that the subjects followed the text in smooth pursuit. The RSVP format resulted in the least number of eye movements, but far from eliminated them (Figure 19-22).

DISCUSSION

The discussion will focus on three issues: What was learned about readability on mobile devices based on the experiments, what was learned about evaluating readability based on the experiences, and finally, what is hoped to be learned from readability evaluations on mobile devices in the future.

Just as in most previous evaluations, the focus has been on improving dynamic text presentation in the RSVP format. There is probably something with the RSVP format that attracts the attention of researchers, unfortunately, however, the format has not been found to live up to the expectations. It has repeatedly been found to be less likeable than other formats, even if it is just as efficient. The idea that the RSVP format, by eliminating eye movements, should increase reading speed and reduce cognitive load, does not seem to hold; the experiments point in the opposite direction. In both studies performed on mobile phones, RSVP was far less efficient compared to the other formats. A partial explanation for the poor results in study four is that the experiment only contained one condition using RSVP. To be fair, none of the subjects participating in the experiments had any previous experience of using RSVP, or at least extremely limited, compared to the other formats. There is probably a learning curve for using RSVP. The question is how much of this curve can be seen in the experiments and how much training can be expected for users to put in. A new text presentation format does not really let you do new things; it lets you read things in a new way. For a new text presentation format to rival existing formats it probably must offer an immediate gratification, either in terms of increased readability or something else. RSVP as it is commonly implemented today, does not offer this gratification.

The page format has, in the experiments, repeatedly been found to excel in terms of readability. On the PDA it was understandable since the screen was not much smaller than a pocket book and the Microsoft Reader program was very well designed. On the mobile phone it worked surprisingly well although the reading speed dropped by 25% compared to the preceding experiment (178 vs. 239 wpm), interestingly the drop for word RSVP format was similarly large 24% (135 vs.177 wpm). Since the preceding experiment was not fully balanced, an educated guess would be that the decrease for the page format would have been less if the experiment were properly balanced. Compared to study one, reading speed for the page format only dropped by 9% (217 vs. 239) when used on a mobile phone as opposed to on a PDA. This is interesting since this is exactly the same figure Duchnicky and Kolers (1983) found when reducing window heights from 20 to 4 lines on a CRT screen. Word RSVP on the mobile phone was 30% slower than chunked RSVP on the PDA (192 vs. 135). A decrease in speed is understandable for the page format when used on a smaller screen; that the difference is smaller than 10% is surprising. The RSVP format was not expected to decrease at all in reading speed, as it should not require more space. One reason for the conflicting results is, yet again, probably stemming from the faulty design of study two. It may be the case that the bad results for the buffered RSVP format tainted the results for the chunked RSVP format. The buffered RSVP format was really bad in terms of readability. Since the chunked RSVP format was more similar to the buffered format than the word format, it might have been read slower and received higher task load ratings. This is just speculation, but a follow up experiment with a proper design might show that chunked RSVP actually is preferable to word RSVP on a mobile phone.

As mentioned, evaluating readability is difficult. If nothing else, the experiments presented in this chapter illustrate how easy it is to go astray. To use within-subject repeated-measurement experimental designs that are balanced according to latin-square designs seems like a good methodology. If used properly, it can limit the number of subjects while still controlling the effects of text, format, and order. Used wrong, it can cause misleading results. It is a powerful tool but must be used with caution. In the experiments, the full potential of the model has not been realized. When creating the design, it is important that it is randomized. Graeco-latin-squares made up of presentation formats and texts were used. These were then transposed over presentation order and text to get a randomization. A downside with transposition is that one certain type of format more often than randomly, is read after a second type of format, although it happens in different positions and with different text. The same pattern occurs with the texts. The transposed randomization can, however, be improved. Four special sets of fourth order orthogonal graeco-latin-squares can be combined into a hyper-latin-graeco-square (HGLS) (Box et al., 2005). Four text presentation formats can then be evaluated with 16 subjects in a design where text, format, and order are perfectly randomized. Using such a design might have limited the eventual contamination effects of buffered RSVP on the results of chunked RSVP in study two. It is hard to assess the effects of the ties in the design that was used; it has probably had an impact on the results. The question is how large the error is; a follow up experiment replicating the last study may shed some light on this issue.

In the last study, paging was found to offer better readability than scrolling. This is interesting since scrolling is the text presentation format predominantly used on mobile phones today. The implementation of scrolling on the phone did, however, only scroll line by line. A scrolling format that moved more lines at a time might improve the format. A combination of scrolling and paging would, however, probably be a good idea; the up and down buttons can be used for scrolling whereas the left and right buttons can be used for paging. The least number of interactions (e.g., button presses) needed for scrolling to read a text was approximately 150 as opposed to 50 for the page format; this is, however, not reflected in physical demand which could indicate that the number of interactions required does not have such a large impact on task load as could be expected. The format that resulted in the highest task load ratings was the leading format. The factor for physical demand is very high, especially considering that the format is dynamic and requires very little user interaction. It may very well be the case that the high amount of unnatural eye movements actually resulted in physical strain. This is a clear indication that leading may not be very suitable for extended text presentation although it was efficient in terms of speed and comprehension. A dynamic presentation format that moves the text in a fashion more similar to how we actually read may turn out to be equally efficient, but much less demanding. The most suitable text presentation formats on mobile phones seem to be the page format and RSVP, of these two, RSVP is the format with greatest potential for improvement.

CONCLUSION

As seen in this chapter, any new format wanted to be used for text presentation must conform to how we are used to reading. Regardless of the device used for reading or the format used for text presentation, the physiological and cognitive limits for reading remain the same. With a starting point in our ability to read, it has been seen how readability can be defined and measured. A review of text presentation formats intended for mobile devices have been presented together with results from previous findings. A methodology for evaluating readability based on a latin-square balanced repeated-measurement experimental design was introduced. This was exemplified by three studies where eye movement tracking was used as a tool to explore readability on mobile devices. The results from the three evaluations show that the latin-square design is useful, but must be implemented correctly. The eye movement studies showed that the page format repeatedly was most efficient, both on a PDA and a mobile phone. In fact, using paging on a mobile phone was only about 10% less efficient than using it on a PDA. The RSVP format did not live to the expectations, clearly the elimination of eye movements does neither increase reading speed nor decrease task load. Future evaluations using chunked RSVP rather than word RSVP, or a format that stimulates eye movements rather than suppresses them, may prove fruitful. Leading was found to be efficient on a mobile phone in terms of reading speed, the unnatural eye movements required for reading does, however, seem to induce too much strain to be acceptable. The methods used in the evaluations have been discussed and a further improvement of the latin-square design has been introduced. This chapter has shown why readability on mobile devices is important, how it may be evaluated in an efficient yet reliable manner, and finally pinpointed RSVP as the format with greatest potential for improvement.

REFERENCES

Björnsson, C. H. (1968). *Läsbarhet.* Stockholm: Liber.

Bruijn, O., & Spence, R. (2000). Rapid serial visual presentation: A space-time trade-off in information presentation. In *Proceedings of Advanced Visual Interfaces (AVI2000)* (pp. 189-192).

Bonferroni, C. E. (1935). Il calcolo delle assicurazioni su gruppi di teste. In *Studi in Onore del Professore Salvatore Ortu Carboni* (pp. 13-60). Rome.

Box, G. E. P., Hunter, W. G., & Hunter, S. J. (2005). *Statistics for Experimenters: Design, Innovation, and Discovery* (2nd ed.). New York: John Wiley & Sons.

Castelhano, M. S., & Muter, P. (2001). Optimizing the reading of electronic text using rapid serial visual presentation. *Behaviour & Information Technology*, *20*(4), 237-247.

Cocklin, T. G., Ward, N. J., Chen, H. C., & Juola, J. F. (1984). Factors influencing readability of rapid presented text segments. *Memory & Cognitio*n, *12*(5), 431–442.

Danvall, L. (2004). *Läsning på små skärmar—En studie med ögonrörelseregistreringsteknik*. Optician thesis, Department of Clinical Neuroscience, Karolinska Institutet.

Dillon, A. (1992). Reading from paper versus screens: A critical review of the empirical literature. *Ergonomics*, *35*(10), 1297-1326.

Dillon, A., Richardson, J., & McKnight, C. (1990). The effect of display size and text splitting on reading lengthy text from screen. *Behaviour & Information Technology*, *9*(3), 215-227.

Duchnicky, R. L., & Kolers, P. A. (1983). Readability of text scrolled on video display terminals as a function of window size. *Human Factors*, *25*, 683-692.

Fine, E. M., & Peli, E. (1995). Scrolled and rapid serial visual presentation texts are read at similar rates by the visually impaired. *Journal of Optical Society of Americ*a, *12*(10), 2286-2292.

Forster, K. I. (1970). Visual perception of rapidly presented word sequences of varying complexity. *Perception and psychophysics*, *8*, 215-221.

GSM Association. (2006, October 17). *Universal access—How mobile can bring communications to all.* GSM Association (GSMA) report.

Gould, J. D., Alfaro, L., Finn, R., Haupt, B., & Minuto, A. (1987). Reading from CRT displays can be as fast as reading from paper. *Human Factors*, *29*(5), 497-517.

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In, P. A. Hancock & N. Meshkati (Eds.), *Human mental workload*. North-Holland: Elsevier Science Publishers, B.V.

Jackson, M. D., & McClelland, J. L. (1979). Processing determinants of reading speed. *Journal of experimental psycholog*y, *108*, 151–181.

Juola, J. F., Tiritoglu, A., & Pleunis, J. (1995). Reading text presented on a small display. *Applied Ergonomics*, *26*, 227-229.

Juola, J. F., Ward, N. J., & McNamara, T. (1982). Visual search and reading of rapid serial presentations of letter strings, words, and text. *Journal of Experimental Psychology: General*, *111*, 208-227.

Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review*, *87*(4), 329-354.

Kang, T. J., & Muter, P. (1989). Reading dynamically displayed text. *Behaviour & Information Technology*, *8*(1), 33-42.

Laarni, J. (2002). Searching for optimal methods of presenting dynamic text on different types of screens. In *Proceedings of NordiCHI 2002* (pp. 217-220). New York: ACM Press.

Lundin, K. (2006). *Läsning på mobiltelefon*: *En studie med ögonrörelseteknik*. Optician thesis, Department of Clinical Neuroscience, Karolinska Institutet.

Masson, M. E. J. (1983). Conceptual processing of text during skimming and rapid sequential reading. *Memory and Cognition*, *11*, 262-274.

Mills, C. B., & Weldon, L. J. (1987). Reading text from computer screens. *ACM Computing Surveys*, *19*(4). ACM Press.

Muter, P. (1996). Interface design and optimization of reading of continuous text. In H. van Oostendorp & S. de Mul (Eds.), *Cognitive aspects of electronic text processing.* Norwood, NJ: Ablex.

Muter, P., Kruk, R. S., Buttigieg, M. A., & Kang, T. J. (1988). Reader controlled computerized presentation of text. *Human Factors*, *30*, 473-486.

Muter, P., & Maurutto, P. (1991). Reading and skimming from computer screens and books: The paperless office revisited? *Behaviour & Information Technology*, *10*, 257–266.

Ober, J. (1994). Infra-red reflection technique. In J. Ygge & G. Lennerstrand (Eds.), *Eye movements in reading.* Elsevier Science.

Osborne, D. J., & Holton, D. (1988). Reading from screen vs. paper: There is no difference. *International Journal of Man-Machine Studies*, *28*, 1-9.

Öquist, G. (2004). Enabling embodied text presentation on mobile devices. In *Proceedings of Mobile and Ubiquitous Information Access 2004* (pp. 26-31). Glasgow, Scotland.

Öquist, G., & Goldstein, M. (2003). Towards an improved readability on mobile devices: Evaluating adaptive rapid serial visual presentation. *Interacting with Computers*, *15*(4), 539-558.

Öquist, G., Goldstein, M., & Chincholle, D. (2004). Assessing usability across multiple user interfaces. In A. Seffah & H. Javahery (Eds.), *Multiple user interfaces: Cross-platform applications and context-aware interfaces* (pp. 327-349). New York: John Wiley & Sons.

Öquist, G., Sågvall-Hein, A., Ygge, J., & Goldstein, M. (2004). Eye movement study of reading text on a mobile device using the traditional page and the dynamic RSVP format. In *Proceedings of Mobile HCI 2004* (pp. 108-119). Glasgow, Scotland: Berlin: Springer.

Paulson, L. J., & Goodman, K. S. (2000). Influential studies in eye movement research. *International Reading Association*.

Potter, M. C. (1984). Rapid serial visual presentation (RSVP): A method for studying language processing. In D. E. Kieras & M. A. Just (Eds.), *New methods in reading comprehension research*. Hillsdale, NJ: Erlbaum

Proctor, R. W. & Proctor, J. D. (1997). Sensation and perception. In G. Salvendy (Ed.), *Handbook of human factors and ergonomics* (2nd ed.) (pp. 53-57). New York: Wiley-Interscience.

Rahman, T., & Muter, P. (1999). Designing an interface to optimise reading with small display windows. *Human Factors*, *1*(1), 106-117.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, *124*, 372-422.

Reichle, E. D., Rayner, K., & Pollatsek, A. (2000). Comparing the E-Z reader model to other models of eye movement control in reading. *Cognitive Sciences E-print Archive*. Retrieved from http://cogprints.soton.ac.uk

Reilly, R. (1993). A connectionist framework for modeling eye-movement control in reading. In G. d'Ydewalle & Van Rensbergen (Eds.), *Perception and cognition: Advances in eye movement research* (pp. 191-212). Amsterdam: North Holland.

Robeck, M. C., & Wallace, R. R. (1990). *The psychology of reading: An interdisciplinary approach* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Sahami, M., Mittal, V., Baluja S., & Rowley, H. (2004). The happy searcher: Challenges in web information retrieval. In *Proceedings of the 8th Pacific Rim International Conference on Artificial Intelligence* (pp. 3-12). Springer.

Schneiderman, B. (1982). The future of interactive systems and the emergence of direct manipulation. *Behaviour and Information Technology*, *1*, 237-256.

Shneiderman, B. (1998). *Human-Computer Interaction* (3rd ed.) (pp. 412-414). Addison Wesley Longman, Inc.

Shaywitz, B. A., Lyon, R. G., & Shaywitz, S. E. (2006). The role of functional magnetic resonance imaging in understanding reading and dyslexia. *Developmental Neuropsychology*, *30*(1), 613-632.

Sicheritz, K. (2000). *Applying the rapid serial presentation technique to personal digital assistants*. Master's Thesis, Department of Linguistics, Uppsala University.

Tekfi, C. (1987). Readability formulas: An overview. *Journal of Documentation*, *43*(3), 261-73.

Wickens, C. D. (1992). *Engineering psychology and human performance* (2nd ed.). New York: HarperCollins Publishers Inc.

Wireless Intelligence. (2006, Septmeber 7). *Worldwide cellular connections pass 2.5 billion*. Wireless Intelligence survey.

Ygge, J., Bolzani, R., & Tian, S. (1999). A computer based system for acquisition, recording and analysis of 3D eye movement signals. In C. Pritchard (Ed.), *Transactions IX International Orthoptic Congress* (pp. 91-94). Stockholm.

KEY TERMS

Dynamic Text Presentation: Text presented using formats that dynamically displays the text over time

Leading: A dynamic text presentation format displaying texts by scrolling continuously on a single line

Page Format: A traditional text presentation format displaying texts by using several screen sized pages

Rapid Serial Visual Presentation: A dynamic text presentation format displaying texts one or a few words at a time

Readability: The ease with which the reading process can proceed

Scrolling Format: A traditional text presentation format displaying texts by using a page larger than the screen

Traditional Text Presentation: Text presented using formats based on how text is traditionally presented on paper

Chapter LVII
Did You See That?

Murray Crease
National Research Council of Canada, Canada

Joanna Lumsden
National Research Council of Canada, Canada

ABSTRACT

The development of appropriate lab-based evaluation techniques for mobile technologies requires continued research attention. In particular, experimental design needs to account for the environmental context in which such technologies will ultimately be used. This requires, in part, that relevant environmental distractions be incorporated into evaluations. This chapter reflects on different techniques that were used in three separate lab-based mobile evaluation experiments to present visual distractions to participants and to measure the participants' cognizance of the distractions during the course of mobile evaluations of technology. The different techniques met the different needs of the three studies with respect to the fidelity of the data captured, the impact of acknowledging distractions on the evaluation task, and the typical context of use for the technology being evaluated. The results of the studies showed that the introduction of visual distractions did have an impact on the experimental task and indicate that future work is required in this area.

INTRODUCTION

Recent research has demonstrated the effect of mobility on users' task performance when using mobile technologies (e.g., Brewster, 2002; Brewster, Lumsden, Bell, Hall, & Tasker, 2003; Crossan, Murray-Smith, Brewster, Kelly, & Musizza, 2005; Mizobuchi, Chignell, & Newton, 2005; Mustonen, Olkkonen, & Hakkinen, 2004). These studies reflect an increasing recognition of the need to assess the usability of mobile technologies and applications under *mobile* conditions. In this respect, studies like these demonstrate a significant advancement on previous static evaluations of mobile technologies which highlighted cosmetic usability issues (Kjeldskov & Stage, 2004) but failed to uncover usability issues arising as a result of the environmental and physical context in which the technologies would ultimately be used. That said, with minor exceptions, these studies do not adequately incorporate the environmental context of use. Indeed, as noted by Brewster (2002), further research is required to develop appropriate evaluation techniques for the evaluation of mobile devices in *realistic* situations.

Although one might argue that realistic usability evaluations are only achievable in the field, a recent investigation has shown that there is little or no benefit to undertaking evaluations in the field as opposed to in the lab (Kjeldskov, Skov, Als, & Høegh, 2004). Not only have lab-based mobile evaluations of mobile technologies been found to identify more usability problems (including context-specific problems) than field-based studies, but also the lab environment allows for far greater experimental control and easier data capture than is possible in the field (e.g., Goodman, Brewster, & Gray, 2004; Kjeldskov et al., 2004; Kjeldskov & Stage, 2004). That is not to say that field-based evaluations do not have a significant role to play; rather, they are perhaps more suited to later-stage high-level usability evaluations once the majority of usability defects have been eliminated via a process of *controlled* experimentation.

Having concluded that lab-based studies are a viable means by which to assess the usability of mobile technologies, to meaningfully evaluate the *suitability* of the technologies for a given usage scenario, it is essential that the lab set-up adequately reflect the intended context of use of the technology. Perhaps one of the most obvious contextual constraints imposed on mobile technology by the environment in which it is used is the requirement for a user to be visually cognizant of his or her physical surroundings while using the technology. Although the nature of environmental surroundings and the consequence of failing to remain cognizant of such surroundings will obviously differ from scenario to scenario, this constraint typically means that users cannot devote all of their visual resources to the technology with which they are interacting. This chapter reflects on different techniques that were used in three separate lab-based experiments to present visual distractions to users and to measure the users' cognizance of the distractions during the course of mobile evaluations of technology. The experiments were all part of an ongoing research effort to develop contextually-relevant experimental techniques for lab-based mobile usability and suitability evaluations. We are simultaneously investigating the use of other sensory distractions and mechanisms of mobility in order to support comprehensive lab-based studies, but this work is outside the scope of this chapter.

In the following sections, each of the three experimental set-ups is outlined and our observations as to the effectiveness of each in terms of the inclusion of visual distractions are discussed. Our aim is to contribute to an emerging body of knowledge related to achieving meaningful and effective lab-based mobile evaluations. We hope that the reflection on our experience will help guide researchers and/or at least provoke discussion on the issue and, to this end, this chapter closes with some high-level conclusions from our observations.

CLASSIFYING DISTRACTIONS

There are numerous types of distractions affecting different senses that can influence the use of mobile technology. For the purpose of discussion, we classify such distractions as follows:

- **Passive distractions:** These distract users but require no active response
- **Active distractions:** These require a user to respond or react in some way (the required response varying according to the nature of the distraction)
- **Interfering distractions:** These may be passive or active and they interfere with a user's ability to effectively interact with a mobile device

Notably, any given environmental distraction can fit differently into this classification scheme depending on the given context of use. In a lab-based environment, however, the nature and extent of distractions can be controlled to enable an evaluator to focus on, and measure of the effect of, specific distractions on users' ability to interact with mobile technology; this precise level of control is not an option in the field.

As previously mentioned, the remainder of this chapter reflects on our observations relating to the use of active, *visual* distractions (which in some cases are also interfering distractions) in three separate mobile evaluations.

A TALE OF THREE STUDIES

Before considering in detail the specifics of the set-up of each of the three studies, it is useful to reflect on the commonality between the three experimental designs. In each case, visual distractions were presented in the form of images projected on the walls of the lab around the participants. In all three experiments, participants were required to react to a particular subset of the projected images (typically, these represented hazardous situations relative to the given context of use). A generic experimental server was implemented that managed the projectors—as well as other lab resources—to ensure both that the projections could be quickly and easily incorporated into these (and future) experiments and that the sequences in which the projections were displayed were consistently and reliably repeatable. The manner in which participants were required to react or respond to the subset of distractions differed across all three studies.

Study 1

Study 1 was designed to compare the effectiveness of, and user preference for, two different text input options (speech-based interaction and stylus-based interaction) for use with a multimodal mobile field data entry application for the construction industry (specifically, concrete test technicians). To appropriately reflect the anticipated context of use within our study design, we had to consider the key elements of a construction site that would potentially influence a technician's ability to use one or both of the input techniques. On the basis of field observations and interviews with test technicians, we determined these to be: (a) the typical extent of mobility of a technician while using the application; (b) the auditory environmental distractions surrounding a technician; and (c) the visual or physical environmental distractions surrounding a technician—that is, the need for a technician to be cognizant of his physical safety when on-site. In terms of (a), we forced study participants to be mobile when using the application by requiring them to move between points in the lab to retrieve data for entry into the application (see Figure 1). With respect to (b), we introduced interfering distractions in the form of construction level noise (70dB–100dB) to the lab space via a 7.1 surround sound system while participants completed their data entry tasks; we used a between-groups study, with three groups of participants each exposed to a different noise level to help determine thresholds at which one of the interaction techniques had an advantage over the other (this level of control would not have been possible in the field). Eighteen people participated in this study, 6 per group: ages ranged from 18-45, 10 males and 8 females; one third of the participants were civil engineers, the remainder had a diverse range of backgrounds. The precise details of these aspects of the study design can be found in (Lumsden, Kondratova, & Langton, 2006); the purpose of this chapter is, however, to reflect on (c)—the visual distractions employed as part of the experimental design.

A concrete technician needs to be cognizant, for safety reasons, of his surroundings on a construction site. As such, it was important to evaluate the extent to which, when using the application, a technician's visual resource would likely be engaged with the application such that he could not attend to surrounding dangers—in this context, heavy equipment moving. To reflect the fact that a technician needs to be aware of his surroundings on-site, we used a series of 6 ceiling mounted projectors in the lab to project photographic images around the 4 walls of the lab space (see Figure 1). These images included a series of 'safe' construction site photographs (that is, with no heavy equipment) and one 'danger' photograph (a concrete truck); the (*active distraction*) photographs were displayed in a pre-set sequence with randomized location and duration such that all projectors were simultaneously projecting photographs but only one was projecting the 'danger' photograph at any given point in time. While using the application to enter data, participants were required to be conscious of the images being projected around them (the active visual distractions) and to maintain a mental tally of the number of 'danger' photographs of which they were aware. Participants reported this tally to us at the end of their data input task and the ratio of observed 'danger' photographs to actual 'danger' photographs was considered to be a measure of participant awareness.

Observations

The use of active visual distractions in this way was not only representative of the real world situation in which the technology would be used (that is, distractions required awareness but not necessarily immediate action), but it also allowed us to avoid placing any extra demands on the already limited resources of the mobile device. Furthermore, by simply asking participants to mentally note 'danger' photographs, we did not interfere with the interaction mechanisms themselves that were being evaluated were not interfered with (the option discussed in Study 2 would not have been possible in this case given the use of speech technology). Although this undoubtedly resulted in lower fidelity results than the designs described in Studies 2 and 3, this design led to contextually meaningful results which was the intention.

Since each participant was required to use each of the two interaction techniques in turn, the order in which they were exposed to the input mechanisms was counterbalanced to mitigate against learning effects. This meant, therefore, that each participant was required to go through the procedure of monitoring visual distractions twice. What we found was that participants were significantly more aware of the distractions – and therefore better able to provide an accurate tally of the number of distractions seen—after their second experimental session. We attribute this to their becoming more familiar with the experimental protocol and relaxing into the evaluation activity. We consider this a positive aspect of our experimental set-up's capacity to mimic the real-world in the lab: we suggest that this reflects longer-term use of a mobile device, where the user becomes increasingly comfortable with the device and subsequently reverts to being more cognizant of his surroundings. Extended pre-session training with the input technique would likely ensure that all sessions—rather than just the later sessions—are equally representative. Unlike the mechanism for recording distraction awareness

Figure 1. Experimental participant surrounded by visual distractions (three of six projections can be seen) whilst obtaining and entering data from one of seven positions in the lab

used in Study 3, this study incorporated no visual reminder to participants to remain aware of their surroundings. Although, as is discussed, there are benefits to this reminder, in the context of this particular study, we felt that it was interesting to be able to observe 'natural' behaviour since this better reflects real world circumstances.

Study 2

Study 2 was designed to compare the effectiveness of, and user preference for, two different forms of visual presentation device (glasses-mounted display and fold-down flat panel screen) for a wearable computer. As with Study 1, it was necessary to reflect the anticipated context of use for wearable computers. In this instance, however, the context was not as clearly defined so we worked on the assumption that participants would be using the wearable computer for a location-aware task which would require them to: (a) be mobile; (b) monitor the visual display of the wearable; and, due to their mobility, (c) monitor their immediate surroundings for hazards. With respect to (a) we required the participants to navigate between six clearly numbered locations in the lab; four locations formed the corners of a 16m x 4m rectangle, with the other two locations placed at the mid-points of the long edges of this rectangle. In terms of (b), we displayed sets of three characters on the wearable's display and instructed participants to look for sets where all three characters matched. We introduced (c) by projecting images of country roads which participants were required to monitor for cars (i.e., hazards). This experimental set-up simulates a scenario where a user is required to be mobile while monitoring both location-specific information (presented on the wearable computer's display) and his immediate physical environment for hazards.

The wearable display was used to instruct participants as to the location to which they were to move. Six ceiling mounted projectors were used to project the images the participants were required to monitor. Only the two projectors facing the participants were active at any given time (determined automatically via knowledge of the lab location to which the participant had been instructed to move), and a preset projection sequence was used such that, at most, only one hazard image was ever displayed at any given point in time. Participants were instructed to shout "car" when they noticed a hazard image was displayed and this was recorded manually by the experimenter using a PDA. Three groups of participants undertook the experimental task: the participants in Group 1 were required to monitor projections with a set of six different images used (of which, one was a hazard image); the participants in Group 2 were also required to monitor projections but a set of 18 images was used (of which, 3 were hazard images); and the participants in Group 3 were not required to monitor their environment. Participants in Groups 1 and 2 were asked to respond to hazard images in the same way (i.e., shout "car"). Eighteen people (mainly students) participated in this study, 6 per group. All participants used both visual displays with the order of use counter-balanced across all groups. More details on the design and results of this study can be found elsewhere in this book.

Observations

As with Study 1, the use of visual distractions was representative of the real world. Users were only required to monitor the projectors ahead of them, which is typical of mobile users who need to be aware of hazards or obstacles in their path as they move through their environment. Furthermore, because participants were only required to verbally acknowledge hazards, the hazards had no direct influence on their interaction with the experimental system. This was important since the participant's interaction with the system was limited to pressing a graphical button when the three characters on the wearable's display matched. As with Study 3, a second graphical button could have been added to the display to allow participants to acknowledge projected hazards; it was, however, felt that this would have introduced too many errors where the participants simply pressed the wrong button (thus potentially weakening the data). A small error was introduced to the data due to the delay in the recording of an acknowledgement, but this delay was relatively constant across all participants and was offset by the participants' improved interaction with the experimental application.

In this study, unlike Study 1, the recognition of *individual* hazards was recorded (i.e., higher fidelity data was collected in terms of hazard awareness) which supported more detailed analysis of correlations between hazard recognition and user interactions. This data also highlighted the concept of 'just missed' hazards—that is, hazards that were noticed by participants but only acknowledged after they were no longer projected. A comparison of the data for Groups 1 and 2 suggests that changing the complexity of the distraction images used does not necessarily influence task performance but, as expected, the need to monitor distractions does impact performance (Group 3 performed significantly better).

Study 3

Study 3 was designed to compare the effectiveness of, and user preference for, two different forms of navigational cues (auditory and visual) in a mobile context. As before, the design of the study was intended to mimic a real-life usage scenario—in this case, a situation where a user is navigating through a physical space while looking for something specific. Twenty-four participants (mainly students) were required to navigate through a grid laid out on the lab floor while monitoring visual distractions projected on either side of them. As with Study 2, the participants were split into three groups: Group 1 was exposed to distractions in the form of alphanumeric characters and instructed to look for projections of the character "*"; Group 2 was exposed to the same set of images used for Group 1 in Study 2 and instructed to acknowledge images containing a car; and, as before, Group 3 was not exposed to any distractions. When participants in Groups 1 and 2 noticed a 'hazard' (i.e., a "*" or a car, respectively) they were required to acknowledge it by pressing a button on the experimental interface of the PDA they were using. All participants used both types of navigational cues (exposure was counter-balanced). A full description of the experimental task can be found in Crease & Lau (2004) and more details on the results can be found elsewhere in this book.

Observations

Superficially, in terms of visual distractions, this study was very similar to Study 2; that is, participants were required to be aware of hazards and acknowledge when they were projected. This study did, however, differ from Study 2 in the manner by which the participants' recognition of the visual distractions was recorded. In this case, when a hazard was displayed, participants were required to press a graphical button on the interface of the experimental application with which they were interacting. Clearly, by recording the data directly, any errors introduced via the experimenter when manually recording participants' acknowledgements (as in Study 2) were eliminated. Although this came at the cost of disrupting participants' interaction with the mobile application, we felt that it was justifiable in this instance since participants' interaction with the system was more complicated than simply pressing a button (as was the case in Study 2) and used a more sympathetic input device (a stylus compared to a handheld trackball which proved difficult for users in Study 2)—that is, in relative terms it did not represent significant additional interaction effort.

In this study, participants' hazard acknowledgements were recorded directly onto the mobile device with which they were interacting and the data was then sent wirelessly to a remote server where it was logged. Although this set-up increased the extent of usage of the device's limited resources and was at the expense of valuable bandwidth, it enabled automatic uniform time-stamping across all the data logged by the system which resulted in higher fidelity (and potentially more accurate) data records.

DISCUSSION

The previous section described three studies that utilised visual distractions in different ways. The discussion of each study highlighted some of the key observations specific to the given study. This section discusses the effect of the distractions on the studies' results and the appropriateness of the different methods used to record participant recognition of the distractions. Table 1 summarises the techniques used to incorporate distractions in the three studies.

There are several dimensions along which the studies can be compared, but the two main differentiators are the location of the displayed distractions and the means by which participants were required to acknowledge hazards. In Study 1, the distractions were projected on all four walls around the participant. In Study 2, the distractions were projected on the wall(s) behind the location the participants were currently walking to (i.e., facing the participant). In Study 3, the distractions were projected on the walls on either side of the participants. In all three cases, the location was chosen to best match the anticipated context of use. On a construction site (Study 1) users are required to monitor their environment for potential hazards in all directions. The other studies concerned more generic mobile scenarios where the participants are required to watch for potential hazards in their path (Study 2) or potential points of interest in passing (Study 3).

The means by which participants were expected to acknowledge 'hazards' impacted on the fidelity of the data captured, the likely accuracy of the results, and the impact on the evaluation task. In Study 1, the participants were required to keep a count of the number of 'hazard' projections they noticed in each condition; in Study 2, the participants shouted "car" each time they noticed a 'hazard' projection; and in Study 3, the participants pressed a button on the evaluation interface. While this last approach may seem to be the most appropriate in terms of fidelity and accuracy, it was not the most appropriate for the

Table 1. A summary of the use of distractions in the three studies

	Study 1	**Study 2**	**Study 3**
Location of distractions with respect to participants	On all sides	In front	Either side
How hazards were acknowledged by participants	Counted and total recorded at end of condition	Verbal acknowledgment recorded by experimenter on mobile device	Button press on same mobile device used for evaluation task
Fidelity of results	Low—no information regarding individual hazards	High—timings of individual acknowledgments recorded	High—timings of individual acknowledgments recorded
Scope for error	Some—participants may lose count of number of hazards	Some—time delay introduced through manual recording by third party	Little
Impact on evaluation task	Low—slight increase in cognitive load but no direct impact on interaction	Low—parallel modality used to acknowledge hazard	Some—additional button press required but increase in complexity minimal
Evaluation results	Participants became better at observing 'hazards' in second condition	Monitoring environment for 'hazards' impacted task performance	Subjective workload experienced significantly increased when monitoring environment

other studies. In Study 3, the participants were required to continuously interact with the evaluation application running on an iPAQ. They were required to enter their location every time they moved (by pressing the appropriate location on a map) and, in one condition, they were required to press an on-screen button to view the direction they were to move in. Because of the number of interactions required —and their short, atomic nature—an extra button press to acknowledge a 'hazard' did not greatly effect the participants' interaction. In Study 2, however, the participants were only required to press one button when the characters displayed matched as the evaluator managed the recording of their location. To add a second button to the evaluation interface would have necessitated the participants moving the cursor using an unfamiliar input device (a handheld trackball), potentially leading to input errors and participant frustration. In Study 1, not only was the participants' interaction with the device (entering values into a text field) longer in duration than a simple button press, but it was the focus of the evaluation. The addition of a button press to acknowledge a 'hazard' would therefore have significantly impacted on the results of the evaluation.

The approach taken in Study 2 retained the data fidelity of the technique used in Study 3 but did introduce a slight risk of error due to the additional human element required (the experimenter recording the shout) and the slight time delay. The risk of these factors significantly impacting the results is negligible, however, as the risk of the participant pressing the wrong button is present in Study 3 and, similarly, the additional delay is compensated for by the participants' use of parallel interaction modalities (speech in addition to the interaction with the wearable computer) as opposed to the use of a single interaction modality to both record a 'hazard' and perform the experimental task. This approach, however, was not suitable for Study 1 due to both the evaluation context (the volume in the lab required the use of ear defenders making it impossible for the experimenter to hear a shout) and task (voice input). While an alternative means to communicate an acknowledgement to the experimenter (e.g., a nod or a wave) could have been used, such an approach would have been both more intrusive to the participants and more ambiguous to the experimenter.

Studies 2 and 3 both compared the impact of the inclusion of visual distractions on the experimental task. The distractions impacted both the task performance (Study 2) and the subjective workload experienced by the participants (Study 3). Study 1 did not compare the effect of visual distractions but did find that participants were better able to recognise 'hazards' with practice. This implies that care must be taken when incorporating distractions to ensure that participants are given sufficient practice at recognising 'hazards' to ensure that no training effect is experienced. The higher fidelity results in Studies 2 and 3 led to the creation of the concept of a "just missed hazard"—a hazard projection that a participant correctly recognised but did not record until it was no longer visible. In Study 1, however, such a hazard would not be differentiated from a correctly recognised hazard. It may be that the timeliness of recognising a hazard is unimportant but, if the experimental context demands that it is, then the approach taken in Study 1 is inappropriate.

CONCLUSION

This chapter has discussed the use of visual distractions in lab-based evaluations of mobile technology. Three studies that used different techniques to present and acknowledge distractions were discussed. These studies have demonstrated that it is both practical and desirable to include visual distractions in the evaluations of mobile technology. Some care, however, is required to ensure that both the presentation of the distractions and the means by which participants acknowledge their presence is appropriate with respect to the intended context of use, the interaction technique used by the evaluation software, and the data fidelity required. If this balance is achieved, our early results indicate that visual distractions should be considered when designing an evaluation of mobile technology.

REFERENCES

Brewster, S. (2002). Overcoming the lack of screen space on mobile computers. *Personal and Ubiquitous Computing, 6*, 188 - 205.

Brewster, S., Lumsden, J., Bell, M., Hall, M., & Tasker, S. (2003, April 5-10). Multimodal 'eyes-free' interaction techniques for mobile devices. In *Proceedings of CHI'2003Human Factors in Computing Systems* (pp. 473 - 480). Ft Lauderdale, USA.

Crease, M., & Lau, T. Y. (2004, September 13-16). Evaluating the use of sound as a navigational aid on a mobile device. In *Proceedings of the Workshop on Mobile HCI and Sound 6th International Conference on Human Computer Interaction with Mobile Devices and Services,* (*Mobile HCI'2004*) (pp. 19-24). Glasgow, Scotland.

Crossan, A., Murray-Smith, R., Brewster, S., Kelly, J., & Musizza, B. (2005). Gait phase effects in mobile interaction. In *Proceedings of Human Factors in Computing Systems CHI'05 Extended Abstracts* (pp. 1312 - 1315). Portland, Oregon, USA.

Goodman, J., Brewster, S., & Gray, P. (2004). Using field experiments to evaluate mobile guides. In *Proceedings of HCI in Mobile Guides Workshop—6th International Symposium on Human Computer Interaction with Mobile Devices and Services MobileHCI'04*, Glasgow, Scotland.

Kjeldskov, J., Skov, M. B., Als, B. S., & Høegh, R. T. (2004, September 13-16). Is it worth the hassle? Exploring the added value of evaluating the usability of context-aware mobile systems in the field. In *Proceedings of the 6th International Symposium on Mobile Human-Computer Interaction MobileHCI'04* (pp. 61 - 73). Glasgow, Scotland.

Kjeldskov, J., & Stage, J. (2004). New techniques for usability evaluation of mobile systems. *International Journal of Human Computer Studies (IJHCS), 60*(5-6), 599 - 620.

Lumsden, J., Kondratova, I., & Langton, N. (2006). Bringing a construction site into the lab: A context-relevant lab-based evaluation of a multimodal mobile application. In *Proceedings of the 1st International Workshop on Multimodal and Pervasive Services MAPS'2006*, Lyon, France.

Mizobuchi, S., Chignell, M., & Newton, D. (2005). Mobile text entry: Relationship between walking speed and text input task difficulty. In *Proceedings of the 7th International Symposium on Human Computer Interaction with Mobile Devices and Services MobileHCI'04* (pp. 122 - 128). Salzburg, Austria.

Mustonen, T., Olkkonen, M., & Hakkinen, J. (2004). Examining mobile phone text legibility while walking. In *Proceedings of the Human Factors in Computing Systems CHI'2004- Extended Abstracts* (pp. 1243 - 1246). Vienna, Austria.

KEY TERMS

Active Distractions: Distractions incorporated within a lab-based experiment which require a user to respond or react in some way (the required response varying according to the nature of the distraction).

Context: The intended context of use for a mobile application.

Field Study: The evaluation of a mobile application that takes place in the actual context of use. The advantage of such evaluations is that problems that only arise in the particular context will be detected. The disadvantages of such evaluations include difficulties in controlling the environment and capturing evaluation data.

Interfering Distractions: Distractions incorporated within a lab-based experiment which may be passive or active and they interfere with a user's ability to effectively interact with a mobile device.

Lab Evaluation: The evaluation of a mobile application that takes place in a laboratory. The advantages of such evaluations include ease of controlling the environment and data capture. Disadvantages include difficulties in creating an appropriately realistic evaluation setting.

Passive Distractions: Distractions incorporated within a lab-based experiment which distract users but require no active response.

Visual Distraction: Distractions that require participants' visual attention.

Chapter LVIII
A Field Laboratory for Evaluating in Situ

Rune T. Høegh
Aalborg University, Denmark

Jesper Kjeldskov
Aalborg University, Denmark

Mikael B. Skov
Aalborg University, Denmark

Jan Stage
Aalborg University, Denmark

ABSTRACT

Evaluating mobile technologies "in the real world" is hard. It is challenging to capture key situations of use, hard to apply established techniques such as observation and "thinking aloud," and it is complicated to collect data of an acceptable quality. In response to these challenges, a "field laboratory" has been developed for evaluating mobile technologies in situ. Facilitating high-quality data collection as well as unobstructed user interaction, the field laboratory allows a small wireless camera to be attached to a mobile device, capturing a close-up image of the screen and buttons. This chapter describes the iterative development of the field laboratory over 4 years of evaluating several mobile systems in field settings. It leads to a description of the current setup and how it is used, and explains the rationales for key decisions on technology and form factors made throughout its development.

INTRODUCTION

Studying peoples' use of technology is a key activity within the research field of human-computer interaction (HCI), providing software developers with invaluable information about the usability and usefulness of their systems at different stages of the process from conceptual design to a final implementation. Traditionally, such studies have taken place in dedicated "usability laboratories" where users' interaction with computer systems can be observed in a controlled experimental setting providing video and audio data of very high quality. Studying the usability of mobile technologies, however, raises new questions and concerns. Mobile systems are typically used in highly dynamic contexts involving a close interaction between people, systems, and their surroundings. Therefore, studying mobile technology use in situ seems like an appealing or even indispensable approach—rather than trying to recreate the use situation realistically in a laboratory. However, studying mobile technology usability "in the real world" is difficult. It is difficult to capture key situations of use, apply established usability techniques such as observation and "thinking aloud" without interfering with the situation, and it is complicated to collect data of an acceptable quality.

In response to some of these challenges, our stationary usability laboratory at Aalborg University's Department of Computer Science has been extended with a mobile counterpart, the field laboratory, which can be taken into the field when studying mobile system use and usability. Facilitating high-quality data collection as well as unobstructed user interaction, the field laboratory allows a small wireless camera to be attached to the mobile device, capturing a close-up image of the screen and buttons while a third-person view is captured by a handheld camcorder.

The purpose of this chapter is to communicate the experiences with developing and using the field laboratory for evaluating mobile technology use and usability in situ by taking the readers through 4 years of major iterations leading to its current configuration. By doing this, the aim is to make practitioners, researchers, and designers of mobile technologies able to set up and use their own field laboratories for evaluating mobile systems in situ. The aim is also to inspire further development of even better field laboratory setups facilitating better, easier, faster, and cheaper use and usability data collection in the field. It is not the purpose of this chapter to discuss the relation between evaluating in the field or in the lab. The point of departure is taken in the assumption that you have decided to evaluate in the field and focus on how you can collect high quality data while out there. It is also not the aim to present or discuss findings about the usability of the specific systems that have been evaluated with the field laboratory (these can be found elsewhere). Instead, the purpose of mentioning these studies here is to illustrate how they functioned as vehicles for iterating on the field laboratory's configuration.

The chapter begins with a short summary of related work motivating the development of techniques for improving evaluation data collection in the field. Three iterations of developing the authors' own field laboratory are then described. For each of these iterations, there are descriptions of the initial motivations and aims, the corresponding configuration of equipment, an example evaluation where it was used, and the pros and cons identified. The next iteration then describes how the field laboratory configuration was modified accordingly, and what was learned from using it in practice. Finally, the current setup is described, some future trends within this area of research are outlined, and the work presented in the chapter is concluded on.

BACKGROUND

In the proceedings of the first workshop on Human-Computer Interaction (HCI) for Mobile Devices in 1998, researchers and practitioners were encouraged to investigate further into the criteria, methods, and data collection techniques for usability evaluation of mobile systems (Johnson, 1998). Of specific concerns, it was stated that traditional usability laboratory setups would not adequately be able to simulate the context surrounding the use of mobile systems and that evaluation techniques and data collection methods such as think-aloud, video recording, or observations would be extremely difficult in natural settings. These concerns have since been confirmed through a number of studies such as (Brewster, 2002; Esbjörnsson, Juhlin, & Östergren, 2003; Pascoe, Ryan, & Morse, 2000).

In 2003, a literature study revealed that 41% of mobile HCI research involved evaluation

(Kjeldskov & Graham, 2003). However, even though evaluations of mobile systems were clearly prevalent, only 19% of these evaluations were carried out in the field while 71% were carried out in laboratory settings. Although the issue of how to study and evaluate mobile technology use and usability in the field has since received increased attention, no established set of usability evaluation methods and data collection techniques yet exists for field evaluations.

The research into field-based evaluations of user interfaces for mobile technologies can be divided into two overall categories of equal importance. The first category focuses on the methodological challenges of adapting traditional usability evaluation methods such as the use of the think-aloud protocol, as well as developing new ones, to suit the challenges and prospects of evaluating mobile user interfaces in the field. The second category focuses on the practical challenges of improving existing techniques for data collection in field settings and developing new ones. In this chapter, the latter is focused on: how to facilitate data collection better when evaluating user interface design for mobile technologies in the field.

One of the primary sources of data when evaluating the usability of an IT system is video and audio recordings of use depicting the system, the users' interaction with it, and the context in which this takes place. When evaluating in the field, the primary challenge of data collection is that these recordings can be very hard to make at a sufficient level of quality. Video filming evaluation sessions in the field with a handheld camcorder is seemingly an attractive approach because it is cheap and easy (Figure 1, left). However, while suitable for capturing the overall use context of a field evaluation, capturing good close-up views of mobile device screen, buttons, and user interaction can be quite difficult while moving (Kjeldskov et al., 2005). Furthermore, filming a good overview of a use situation with a handheld camcorder requires a bit of distance, while obtaining good close-ups and good sound requires that the cameraman stay relatively close to the test subject and interviewer. The latter often results in the so-called "bodyguard effect" (Figure 1, right) where the test subject is practically isolated from other people in their surroundings, hence questioning the value of going into the field in the first place (Kjeldskov & Stage, 2004).

Within the "practical" category of improving data collection techniques for evaluations in the field, three specific approaches are particularly worth mentioning.

One approach has aimed at obtaining field data in a non-intrusive way through automatic logging of user interaction for later analysis. Through logging, researchers can accurately record a user's interaction with a system, such as clicks or keyboard entries, or even record the entire graphical user interface of the software being evaluated. One of the advantages of logging is that it does not necessarily require the presence of a test monitor, and involves a minimum of interference with the user's context. This makes logging particularly useful for longitudinal studies of mobile technology usability. Logging is also an

Figure 1. Usability evaluations of mobile technologies in the field using a handheld camcorder and note taking for data collection

efficient method to obtain data in a cost effective way from a large population of users. One of the drawbacks of logging is that it does not usually record any information outside the mobile device. It provides no record of, for example, the physical surroundings of the user, and it does not record so-called "near-interactions" where, for example, the user fails to interact with the system (Waterson, Landay, & Matthews, 2002). Another quite significant limitation is that logging usually requires installation of dedicated software on the device being evaluated. This is not only cumbersome but also, sometimes simply not possible. While highly suitable for generating large amounts of data for quantitative studies, logging does usually not provide good data for qualitative studies. A way of overcoming this limitation could be to combine automatic logging with, for example, video and audio recordings, interviews, and so forth.

Another approach has aimed at bringing traditional laboratory setups into the field by means of a "portable usability laboratory" or "lab-in-a-box" (Kimber, Georgievski, & Sharda, 2005; Winters et al., 2001). The advantage of a portable laboratory is that it allows rich data to be collected using high quality equipment. Not being truly mobile, portable usability laboratories are, however, best used in field settings where the user remains semi-mobile within a delimited spatial area for a period of time—for example, in a restaurant or on the bridge of a ship. Other drawbacks for this approach are that the equipment is often cumbersome to transport and setup and may be intrusive in the context (Rowley, 1994). Setting up large amounts of video and audio recording equipment in the field may also cause users and surrounding people to act differently, which, in essence, stand diametrically opposed to the purpose of evaluating usability in the field. As a final downside, it may be difficult to record video of users' interaction with a mobile device with standard camera equipment.

Taking its offset in the challenges of using portable laboratories in the field, a third approach has been aimed at developing more compact and mobile usability laboratory facilities that are able to record high quality video data from various sources in an un-intrusive way. Different configurations of such mobile usability laboratories have been described in recent literature (e.g., Betiol & Cybis, 2005; Kaikkonen, Kallio, Kekäläinen, Kankainen, & Cankar, 2005; Roto et al., 2004) and demonstrated at leading conferences within the field (e.g., Nyyssönen, Roto, & Kaikkonen, 2002). The typical setup of a mobile laboratory makes use of a mini camera that can be attached to a mobile device for a good close up of the screen and user interaction. In some setups, such as the one proposed by Roto et al. (2004), additional cameras are used to capture views of the evaluation context. Images from these cameras are then mixed and recorded for later playback during analysis. While mini-camera approaches like this are highly promising—not only in field evaluations but also in laboratory settings—experiences from the deployment of mobile usability laboratories in the field also point out a series of challenges. Some of the issues relate to the quality of video and audio recordings when using wireless equipment, and how to best record multiple video sources and audio in sync. Other issues relate to battery lifetime and the weight of the equipment having to be carried around during the evaluation sessions.

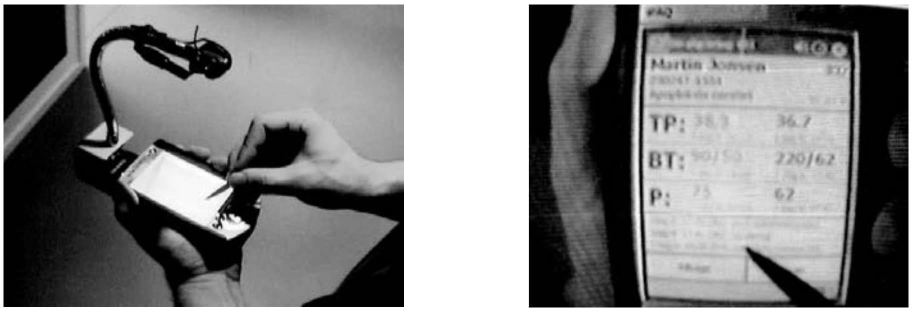
In the following sections, it will be outlined how these and other challenges have been dealt with through three iterations of setting up and using a field laboratory for in situ evaluations of mobile technologies

CLOSE-UP VIDEO AND IMPROVED SOUND

Motivated by the challenges of capturing high-quality video data during usability evaluations in the field described in the literature and experienced in a series of evaluations carried out between 2002 and 2003, it was decided to develop a portable configuration of audio and video equipment that could be carried by the test subject and an observer during a field evaluation.

The primary focus for the first version of the "field laboratory" was to enable close-up recording of the mobile device screen and user interaction.

Figure 2. PDA on camera-mount allowing for close-up view of screen and user interaction



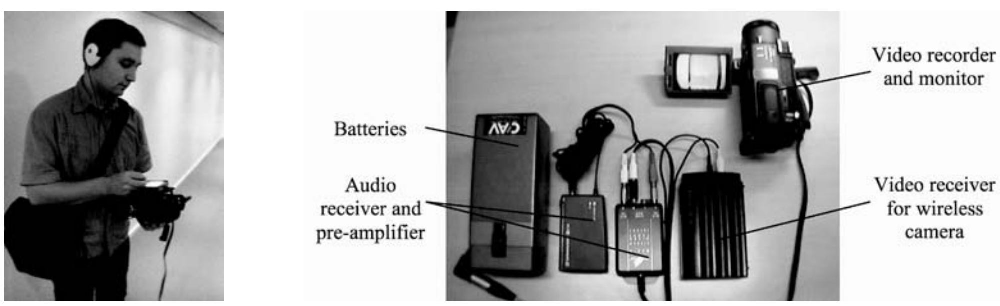
Inspired by commercially available products, such as the "mobile device camera" from Noldus (Noldus, 2005), a small camera-mount on which a mobile phone or PDA could be mounted with Velcro (Figure 2, left) was constructed. The camera-mount contained a wireless camera mounted on a flexible "gooseneck" as well as a 9v battery-supply. This allowed the capturing of a detailed close-up view of the mobile device in colour (Figure 2, right) and the recording of this throughout the whole evaluation. Apart from recording close-up video of the mobile device, what was also wanted was to improve the sound quality of the data recordings to minimize ambient noise and ensure capturing all utterances made by the test subject and the interviewer. For this, the camera on the mobile device was combined with an off-the-shelf professional wireless microphone from Sennheiser; a lapel microphone with a belt-pack transmitter worn by the test subject and a belt-pack receiver carried by the observer.

Video from the camera on the mobile device and audio from the lapel microphone is transmitted wirelessly to receivers and recording equipment carried by an observer (Figure 3). In the observer's bag, the video and audio signals are recorded on a portable DV recorder, for example, a camcorder, set up to record from an external source. During the evaluation, the observer can monitor the user's interaction with the mobile device on a small LCD screen and monitor the sound through earphones.

Using the First Field Laboratory in Practice

The first version of the field laboratory described, was used for an evaluation of a mobile information system in situ in 2003/2004 (Kjeldskov, Skov, Als, & Høegh, 2004). The evaluation focused on the use of a mobile, context-aware, electronic patient record system by nurses and doctors at a large

Figure 3. Observer (left) carrying and operating portable audio/video equipment (right) for capturing close-up view of screen and user interaction



regional hospital in Denmark. Six test subjects (all females) aged between 25 and 55 years participated in the field evaluation. They were all trained nurses with 1-9 years of professional experience.

Due to the real-life nature of the study, the field evaluation did not involve any researcher control in the form of task assignments but was structured exclusively by the work activities of the nurses. The studied work activities were highly mobile, and involved interaction with assigned patients in different wards (i.e., collecting and reporting scheduled measurements) and moving back and forth between different rooms and hallways. As in a standard usability evaluation, the test subjects were given a brief instruction to the mobile system being evaluated and were encouraged to think aloud when possible. Each evaluation session lasted 15 minutes on average and involved three people. One nurse used the system for carrying out her work activities. One researcher acted as interviewer and asked questions for clarification while in the hallway. A second researcher operated the field laboratory. In addition, each session involved a number of hospitalized patients in their beds. For ethical reasons, the hospitalized patients were not filmed. In order to be able to include a suitable number of different nurses as test subjects, the field evaluation took place over 2 days.

Lessons Learned From Using Field Lab #1

The field evaluation at the hospital highlighted a series of the challenges related to evaluating mobile technologies in situ. It was highly time consuming and complex to plan and execute the study, and it was difficult to capture key situations of use. However, in relation to data collection, the camera on the mobile device gave the provision of high-quality close-up views of the nurses' interaction with the system being evaluated, while at the same time, allowing them to move around freely in the environment and focus on their work. The use of a professional wireless microphone supplemented the video close-up recordings with a clear audio track capturing all the nurses' utterances as well as enough ambient sounds to give a sense of context. During the later analysis phase, these video and audio recordings were invaluable sources of data for identifying usability problems and suggesting opportunities for redesign. The video track allowed the sight of exactly which parts of the system were perceived as problematic and where the nurses had problems with operating the interface. The audio track allowed the nurses' comments about their interaction with the system to be heard and provided the context of use. When evaluating mobile technologies in a laboratory, this kind of data is very much standard. The first version of the field laboratory made it possible to capture the same kind of data in situ as well. It was lightweight and it was relatively easy to operate.

On the downside, the first version of the field laboratory also had a number of limitations. First of all, the video recording only contained the close-up view of the mobile device and the user interactions taking place within 5-6 cm of the screen. It did not capture the users or their surroundings. During the data analysis phase, this proved to be very problematic at times where the use context was significant for understanding what the user was trying to do with the system. It was also hard to tell from the video track when the users were looking at the screen of the mobile device and when they focused elsewhere during the evaluation. Although the audio track did provide some information about context and the focus of the users, this information was often partial, ambiguous, and not conclusive. Secondly, the audio track only captured the voice of the interviewer if he or she was standing close to the test subject (who was wearing the microphone). In a stationary evaluation setup, this would usually not be a problem because the interviewer and test subject will be seated close to each other. However, when evaluating in the field, it is most likely that interviewer and test subject will sometimes be physically separated by enough distance for directional microphones not to be able to pick up the voice of them both. In the field evaluation at the hospital, this was often the case simply because the nurses were sometimes hard to keep up with by the interviewer and because the interviewer would sometimes have to stand back a bit in order not to interfere with the nurses' work tasks (i.e., attending to patients in bed). Thirdly, the mini camera was far from perfect. Although it was considerably smaller than commercially available alternatives, the gooseneck camera-mount clearly influenced the form factor of the mobile device being evaluated. It was too heavy and made it impossible for users to hold the device the way they would usually do.

SMALL CAMERAS AND MULTIPLE VIDEO SOURCES

On basis of the lessons learned from the field evaluation at the hospital, it was set out for the field laboratory to be improved in three ways. Firstly, it was wanted that the influence of the wireless camera attached to the mobile device being evaluated be reduced. It was wanted that the size and weight of the camera be minimized and that it be made more flexible for use with different types and sizes of mobile devices. Secondly, it was wanted to facilitate data recording from multiple sources of video allowing the capturing of close-up views of the mobile device, close-up views of the user, third person views of the user in context, and first person views of the surroundings as seen in, for example, Roto et al. (2004). Thirdly, it was wanted that audio from multiple sources possible to be captured independently (e.g., the test subject and the interviewer).

Minimizing the size and weight of the camera on the mobile device turned out to be surprisingly simple, while at the same time also increasing its flexibility. The solution was to simply strap the camera house on to a small plastic clamp with a flexible piece of plastic and a few cable strips. All items necessary to produce the wireless "camera-clamp" were purchased from a local hardware store for less than $20. The clamp made it possible to mount the camera on almost any mobile device without interfering with its form factor (Figure 4). The 9v battery powering the camera was simply attached to the mobile device with double-sided tape, wherever it would interfere the least with the user's grip of it. Using the same approach, other variations of the camera-clamp were created. One was also clipped-on to the mobile device but faced the camera towards the user (Figure 4, right). Another one was designed to sit on the user's ear (like a Bluetooth headset), capturing a first-person view of the surroundings. These additional wireless cameras allowed us to capture video data from multiple sources in parallel.

In order to capture a third-person view of the evaluation session, it was decided that the observer be equipped with a handheld video camcorder. For better audio capture, a second wireless lapel microphone for the interviewer was added.

While reducing the size and weight of equipment carried by the test subjects (even though more cameras were added), the addition of more cameras and microphones significantly increased the equipment necessary to be carried and oper-

Figure 4. Lightweight camera-clamps attached to mobile devices

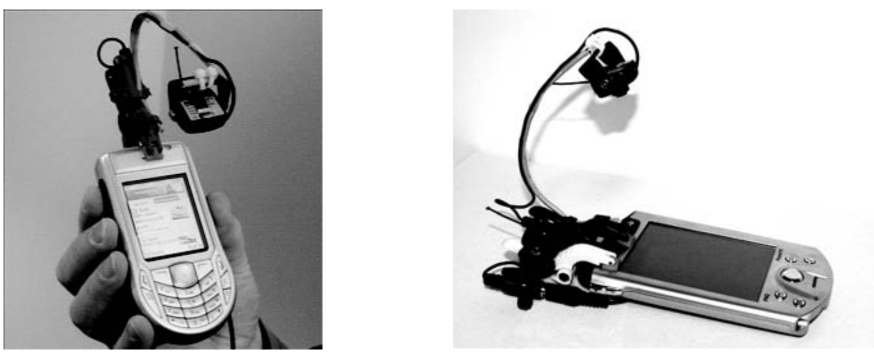
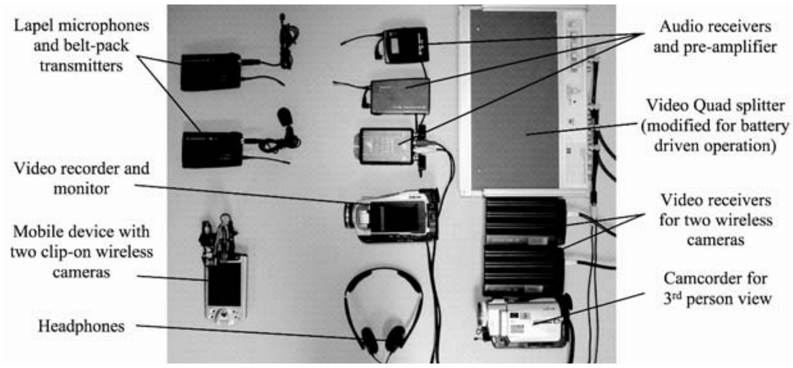


Figure 5. Equipment used for the second field laboratory (configured for two wireless cameras). Batteries and power regulators are not shown



ated by the observer (Figure 5). The additional lapel microphone required an additional belt-pack receiver. For each additional wireless camera, another video receiver and 12v battery had to be added. In order to include the video signal from three to four different sources in one composite video recording, some sort of battery driven video mixing had to be included as well. For this purpose, a stationary Panasonic WJ-MS 424 Quad display unit was modified to run on batteries. In order to minimize the number of different batteries in use and avoid batteries running flat at different times, a power supply, which could power all equipment from the same 12v battery source (apart from the camcorders which ran on their own batteries), was custom-built.

Using the Second Field Laboratory in Practice

The second field laboratory setup described was used in pilot studies preparing for a large-scale evaluation of a mobile information system in situ in 2005. The aim was to facilitate data collection about the use and usability of a context-aware mobile Web site used by pairs of friends while socializing "out on the town." Hence, it was important to document both peoples' interactions with the device, with each other, and with their physical surroundings.

Lessons Learned From Using Field Lab #2

The pilot field evaluations in the city centre of Melbourne once again highlighted the complexity of evaluating mobile technologies in situ. However, this time it was clear that more was gotten out of the efforts to move from the laboratory into the field. The second version of the field laboratory made it possible to capture multiple video and audio sources in situ. As in the evaluations at the hospital, test subjects could move around relatively freely, and were undisturbed by the cameraman who could easily keep a distance of 5-8 meters while still capturing good images and sound. As aimed for, the second field laboratory provided rich data of high quality capturing both detailed views of the users, their interaction with the device, and their surroundings from several perspectives (Figure 6). During the later analysis phase, especially, the third-person view of the users in context provided an invaluable resource for contextualizing peoples' verbal utterances and their interaction with the system. Unlike the early evaluations in the field, where only a handheld camcorder was being used, the field laboratory allowed the cameraman to remain focused on the surroundings rather than having to zoom back and forwards between a third-person view and a close-up view of the mobile device. The use of two microphones resulted in a stereo audio track,

Figure 6. Example recording with multiple video sources

which very clearly captured all utterances by test subjects and interviewer. Recording two separate audio tracks made it easy to separate between utterances made by different people during playback for analysis. It also made it possible to make post-evaluation adjustments of the relative levels of peoples' voices.

On the downside, however, it only took two pilot sessions in the field before realizing that the current setup of the field laboratory had a series of fundamental problems and needed to be modified. While it was possible to capture great data like never before, the cost of this was very high in terms of battery life, weight, and complexity of operating the equipment needed. All the field laboratory equipment depicted in figure 5, as well as the necessary batteries and power supply regulators, were able to be fitted into a large laptop bag with internal cabling. However, the total weight of the bag exceeded 10 kg, which turned out to be physically challenging for the cameraman to carry for more than a few hours. At the same time, the modified Quad display splitter and the video receivers ran the battery-pack of four 12v motorcycle batteries flat in less than 1.5 hours. In effect, this made back-to-back evaluation sessions impossible without recharging or carrying extra batteries into the field as well! While running all equipment on the same 12v power supply reduced the task of monitoring and replacing a lot of individual batteries for, for example, the audio receivers, it was also found that the power regulators needed for doing this introduced noticeable noise to the audio recordings. Finally, the amount of equipment and the number of different video and audio sources made it highly complex for one person to operate the field laboratory in the (already) stressful conditions of an evaluation in situ.

On top of these problems, the number of different wireless technologies involved at this stage also resulted in problems with radio interference between equipment operating on the same or close frequencies. While there were no problems whatsoever with the professional wireless microphones, wireless video from multiple cameras turned out to be problematic. Camera signals sometimes interfered with each other, as well as with the wireless capabilities of the mobile device being evaluated (WLAN and Bluetooth). In fact, using more than one wireless camera at a time sometimes completely disrupted the PDAs WLAN connection, making parts of the evaluation impossible to carry out. At other times, the use of Bluetooth significantly distorted the images from some wireless cameras. Dealing with the problem of radio interference was quite a challenge. While being able to modify the use of wireless technologies to some extent during the evaluations to avoid problematic combinations of Bluetooth, WLAN, and the wireless cameras, evaluating in the field, of course, made it impossible to control *other peoples*' nearby use of wireless technologies, which sometimes interfered with the equipment.

On the bright side, however, revisiting the field recordings quickly made it evident that collecting data from four independent video sources was not necessary in order to get a sufficient view of users, use, and context. The only sources that any significant use was made of during the analysis of the evaluation sessions were the close-up view of the device and the third-person view of users and context. Hence, the equipment could be reduced.

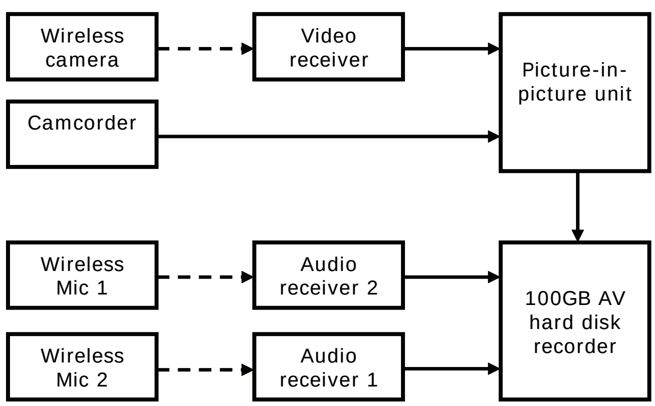
MINIMIZING EQUIPMENT AND INCREASING BATTERY LIFETIME

Informed by the lessons learned from the pilot field studies described, some significant changes were made to the field laboratory with the aim of minimizing equipment, reducing weight and complexity, and increasing battery life.

The first major decision was to reduce the number of video sources to two: a wireless camera attached to the mobile device and a handheld camcorder operated by an observer. Reducing the number of wireless cameras limited the issue of radio interference and allowed the making of some significant reductions in the equipment to be carried by the observer. Firstly, the number of video receivers could be reduced correspondingly. Secondly, the battery-hungry Quad display unit was able to be replaced with a much smaller Picture-in-Picture unit running on 12v (drivedata DPIP1). In return, these reductions made it possible to phase out a few heavy power regulators and run the field laboratory for almost four times longer on half the batteries. Replacing the wireless audio receivers with newer and more lightweight models (Sennheiser ew100 G2), an audio preamplifier and noise generating power regulators could be phased out while at the same time improving the sound quality. The portable tape-based DV recorder was also replaced with a smaller and more lightweight 100GB AV hard disk recorder (Archos AV400). The third generation of the field laboratory is configured as schematically depicted in Figure 7.

Video signals from the wireless camera attached to the mobile device are sent to a receiver in a small bag carried by an observer, there they are mixed on-the-fly with a third-person view of the users and captured by the handheld camcorder. Ensuring high-quality sound, users and the interviewer are wearing small directional wireless lapel microphones. Mixed video and sound is recorded digitally on a hard disk recorder in the observer's bag. This configuration of the field laboratory weights approximately 4 kg, measures 26x18x30 cm, and has a battery time of approximately 5 hours on two 12v batteries (Figure 8).

Figure 7. Schematic configuration of the current version of the field laboratory with two video sources and two audio sources recorded in one composite digital file



Using the Third Field Laboratory in Practice

The third field laboratory setup described was used in a large-scale evaluation focusing on the use and usability of a context-aware mobile Web site facilitating sociality in the city centre of Melbourne, Australia (Kjeldskov & Paay, 2005). The field evaluation involved 20 people (grouped in pairs). All pairs of users were familiar with the location at which the evaluation took place and frequently socialized there together (Figure 9).

With the purpose of being true to the real-life qualities of studying mobile technology use in situ, the field evaluations were not structured by tasks in a traditional usability evaluation sense of the term. Instead, the evaluations were structured by a set of overall prompts for use of different parts of the system and a list of corresponding interview questions. The socializing activities studied were highly mobile and involved the users moving between several physical locations in the city; bars, cafés, museums, and so forth. Prior to the evaluation, the users were given a 10 minute introduction to the system and were allowed to familiarize themselves with it for 5-10 minutes. Inspired by the constructive interaction approach to thinking-aloud studies with more than one user, the groups were asked to talk among themselves about their perception of and interaction with the system, interrupted only with questions for clarification. The evaluation sessions each lasted between 45 and 70 minutes and took place over several days.

Lessons Learned from Using Field Lab #3

The field evaluation of the mobile Web site was very successful. With the third iteration of the field laboratory, a very useable and stable solution with a good trade-off between supported data sources, weight, and battery lifetime had been reached. It was possible to capture audio and video sources needed for studying the use and usability of mobile technologies in situ, and even to do so in a quality that matched (or even sometimes superseded) the stationary laboratory. The field laboratory was small, light-weight, relatively simple to operate, and had a battery lifetime allowing for 2-3 evaluation sessions in a row without worrying about recharging (at this point the weakest link was, in fact, the battery lifetime of the PDAs used to run the prototype system). It allowed the observer to effortlessly follow the participants and interviewer from a bit of a distance while filming them and their surroundings with the handheld camcorder. In turn, this allowed the interviewer to focus on the participants' use of the mobile system being evaluated without having to worry about data collection. Figure 10 shows an example of the video data recorded in the field.

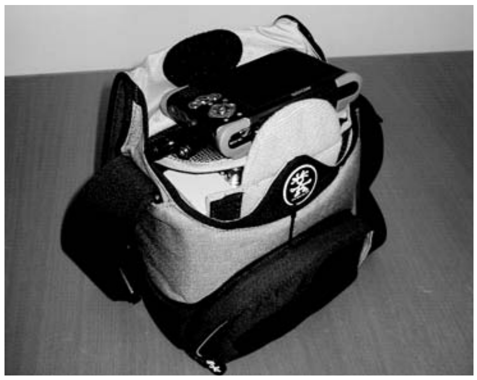
While the third version of the field laboratory was already considerably smaller and lighter than any of the earlier ones, the weight and physical size have since been reduced further through a fourth iteration of reducing cabling, battery supply, and optimizing the use of bag-space (Figure 11). In the most recent design (version 4), the field

Figure 8. The third field laboratory in a medium-sized light-weight camera bag

Figure 9. The field laboratory in action at Federation Square, Melbourne, Australia

Figure 10. Video recording with third-person view of participants and close-up view of PDA. Note that the camera focused on the device screen is turned 90 degrees to optimize use of the Picture-in-Picture view.

Figure 11. The most recent version of the field laboratory weighing only 2 kg and measuring just 18x14x25 cm—containing video and audio receivers, Picture-in-Picture unit, hard disk recorder, and battery.



laboratory has the same specifications for data capture as described above, but now weights only 2 kg and measures only 18x14x25 cm, making it highly mobile and very easy to bring into the field for longer periods of time. Powered by only one 12v battery, this configuration can operate for approximately 2.5 hours before the battery must be swapped with a spare one.

FUTURE TRENDS

The future trends for developing field laboratories for evaluating mobile technology use and usability in situ focus primarily on improving the quality, reliability, and size of the cameras attached to the mobile device. As wireless video technology matures and becomes more widespread, an emergence of cheap high-end wireless video cameras matching the professional standard of the wireless microphones used in the current version of the field laboratory are likely to be seen. Broadcast quality interference-free wireless video technologies exist today, but are still rather expensive and not sufficiently lightweight for our purposes.

Coming from another area of application, new camera technologies are also emerging within the field of video surveillance, which would allow video signals to be transferred digitally via wireless network connections rather than over an analogue radio link. Apart from offering much higher quality and stability, this approach is particularly interesting because it bypasses the use of any analogue video equipment, which is typically quite battery intensive. It also enables the development of field laboratories where all video sources are recorded digitally in separate, time stamped tracks avoiding the down-sampling of Picture-in-Picture and allowing for synchronised playback of multiple camera angles without any loss of quality.

A third emerging way of dealing with the camera problem is to replace it with a software solution that logs screen images from the mobile devices, or replicates them on a laptop or stationary computer via a network connection and then grabs the images from there. However, as discussed earlier in the section about automatic data logging, this approach does not capture the user-interaction with the physical device and situations where, for example, input is not registered by the system. Nevertheless, parallel data logging of the mobile device screen could be a very interesting way of complementing video and audio data captured through wireless cameras and microphones and should be investigated further. In a similar way, capturing video and audio data of user interaction could be an interesting way of enhancing the use of data logging when evaluating mobile technologies in the field.

CONCLUSION

In this chapter, the iterative development of a field laboratory facilitating in situ evaluations of mobile technology use and usability has been described. A series of initial motivations, how we responded to these, and the lessons learned from deploying the field laboratory to a series of evaluations has been described.

It is hard to evaluate mobile technologies in situ. It is difficult to capture key situations of use and it is complicated to collect data of an acceptable quality. However, by means of a field laboratory with small wireless cameras and wireless microphones, it has been shown that it is possible to capture field data about the use and usability of mobile technologies in a quality that matches that of a stationary usability laboratory. Furthermore, it has been shown that field laboratories can be made small, lightweight, and operational for hours before having to recharge batteries. Equipped with a field laboratory as the one described in this chapter, it is believed that researchers and designers will be able to make more and better evaluations of user interfaces for mobile technology in the field.

REFERENCES

Betiol, A. H., & Cybis, W. D. A. (2005). Usability testing of mobile devices: A comparison of three approaches. In *Proceedings of Interact 2005* (pp. 470-481). LNCS: Springer-Verlag.

Brewster, S. (2002). Overcoming the lack of screen space on mobile computers. *Personal and Ubiquitous Computing, 6*(3), 188-205.

Esbjörnsson, M., Juhlin, O., & Östergren, M. (2003). Motorcyclists using Hocman—Field trials on mobile interaction. In *Proceedings of Mobile HCI 2003* (pp. 32-44). LNCS: Springer-Verlag.

Johnson, P. (1998). Usability and mobility; interactions on the move. In *Proceedings of the First Workshop on Human-Computer Interaction with Mobile Devices* (Tech. Rep. G98-1). Glasgow, Scotland: GIST.

Kaikkonen, A., Kallio, T., Kekäläinen, A., Kankainen, A., & Cankar, M. (2005). Usability testing of mobile applications: A comparison between laboratory and field testing. *Journal of Usability Studies, 1*(1), 4-16.

Kimber, J., Georgievski, M., & Sharda, N. (2005). Developing usability testing systems and procedures for mobile tourism services. In *Proceedings of Annual Conference on Information Technology in the Hospital Industry*. Los Angeles: Informational Technology Association.

Kjeldskov, J., & Graham, C. (2003). A review of mobile HCI research methods. In *Proceedings of Mobile HCI 2003* (pp. 317-335). LNCS: Springer-Verlag.

Kjeldskov, J., Graham, C., Pedell, S., Vetere, F., Howard, S., Balbo, S., & Davies, J. (2005). Evaluating the usability of a mobile guide: The influence of location, participants and resources. *Behaviour and Information Technology, 24*(1), 51-65.

Kjeldskov, J., & Paay, J. (2005). Augmenting the city: The design of a context-aware mobile Web site. In *Proceedings of DUX 2005* (pp. 1-7). San Francisco: ACM.

Kjeldskov J., Skov M. B., Als B. S., & Høegh, R. T. (2004). Is it worth the hassle? Exploring the added value of evaluating the usability of context-aware mobile systems in the field. In *Proceedings of Mobile HCI 2004* (pp. 61-73). LNCS: Springer-Verlag.

Kjeldskov, J., & Stage, J. (2004). New techniques for usability evaluation of mobile systems. *International Journal of Human-Computer Studies, 60*, 599-620.

Noldus. (2005). *Mobile device camera*. Retrieved March 2, 2007, from http://www.noldus.com/site/doc200402054

Nyyssönen, T., Roto, V., & Kaikkonen, A. (2002, September). *Mini-camera for usability tests and demonstrations.* Paper presented at demo sessions of Mobile HCI 2002. Pisa, Italy.

Pascoe, J., Ryan, N., & Morse, D. (2000). Using while moving: HCI issues in fieldwork environments. *Transactions on Computer-Human Interaction, 7*(3), 417-437.

Roto, V., Oulasvirta, A., Haikarainen, T., Kuorelahti, J., Lehmuskallio, H., & Nyyssönen, T. (2004). *Examining mobile phone use in the wild with Quasi-experimentation* (HIIT Technical Report 2004-1). Helsinki, Finland: Helsinki Institute for Information Technology.

Rowley, D. (1994). Usability testing in the field: Bring the laboratory to the user. In *Proceedings of CHI 1994* (pp. 252-257). ACM Press.

Waterson, S., Landay, J. A., & Matthews, T. (2002). In the lab or out in the wild: Remote Web usability testing for mobile devices. In *Proceedings of CHI 2002* (pp. 796-797). ACM Press.

Winters, J. M., Story, M. F., Campbell, S., Lemke, M., Danturthi, S., Barr, A., & Rempel, D. M. (2001). Mobile usability lab: A tool for studying medical device accessibility for users with diverse abilities. In *Proceedings of RESNA 27th Annual Conference*.

KEY TERMS

AV Hard Disk Recorder: A video unit that records an external video and audio source directly onto a hard disk in a digital format that can be played back on a computer. The video recording is typically compressed when it is recorded resulting in manageable file sizes.

Camera-Clamp: A tiny camera that can be clipped on to a mobile device such as a PDA or a mobile phone. Camera-clamps can be either cabled or wireless. The latter require a battery supply and a video receiver.

Close-Up View: A video recording of the screen and buttons on a mobile device such as a PDA or mobile phone, during an evaluation of use or usability; usually captured with a mobile devices camera attached to the device.

Field Laboratory: A configuration of laboratory equipment, such as video and audio recording devices, put together so that it can be taken into to field for data collection about the use and usability of mobile technologies in situ.

Lapel Microphone: A small microphone that can be clipped on to a person's collar or revere. The microphone is usually connected to a small transmitter that can be carried in a pocket or clipped on to the belt.

Picture-in-Picture Unit: A video unit that inserts a video image over a part of another one. The inserted video image is rescaled and thus loses a bit of quality in the process.

Quad Display Unit: A video unit that merges four different video signals into one composite signal. All four video images are rescaled and thus lose a bit of quality in the process.

Third Person View: A video recording of the user(s) of a mobile device and their immediate surroundings during an evaluation of use or usability from the perspective of a third person observing from a distance.

Chapter LVIX
Field Evaluation of Collaborative Mobile Applications

Adrian Stoica
University of Patras, Greece

Georgios Fiotakis
University of Patras, Greece

Dimitrios Raptis
University of Patras, Greece

Ioanna Papadimitriou
University of Patras, Greece

Vassilis Komis
University of Patras, Greece

Nikolaos Avouris
University of Patras, Greece

ABSTRACT

This chapter presents a usability evaluation method for context aware mobile applications deployed in semi-public spaces that involve collaboration among groups of users. After reviewing the prominent techniques for collecting data and evaluating mobile applications, a methodology that includes a set of combined techniques for data collection and analysis, suitable for this kind of applications is proposed. To demonstrate its applicability, a case study is described where this methodology has been used. It is argued that the method presented here can be of great help both for researchers that study issues of mobile interaction as well as for practitioners and developers of mobile technology and applications.

INTRODUCTION

Mobile devices are part of many peoples' everyday life, enhancing communication, collaboration, and information access potential. Their vital characteristics of mobility and anywhere connectivity can create new forms of interaction in particular contexts, new applications that cover new needs that emerge, and change the affordances of existing tools/applications.

A case of use of such devices, with particular interest, concerns *public places rich in information* for their visitors, in which mobile technology can provide new services. Examples of such places, are *museums* and other sites of culture (Raptis, Tselios, & Avouris, 2005), *public libraries* (Aittola, Parhi, Vieruaho, & Ojala, 2004; Aittola, Ryhänen, & Ojala, 2003), and *exhibition halls* and *trade fairs* (Fouskas, Pateli, Spinellis, & Virola, 2002). In these places, mobile devices can be used for information collection and exchange, for ad hoc communication with fellow visitors, and for supporting face-to-face interaction.

Usability evaluation of mobile applications is of high importance in order to discover, early enough, the main problems that users may encounter while they are immersed in these environments. Traditional usability evaluation methods used for desktop software cannot be directly applied in these cases since many new aspects need to be taken in consideration, related to mobility and group interaction. Therefore, there is a need either to adapt the existing methods in order to achieve effective usability evaluation of mobile applications or to create new ones. An important issue, that is discussed here, is the *process* and *media* used for recording user behaviour.

Data collection during usability studies is a particularly important issue as many different sources of data may be used. Among them, *video and audio* recordings are invaluable sources for capturing the context of the activity including the users' communication and interaction. It has been reported that in cases of studies that audio and video recordings were lacking, it was not possible to explain why certain behaviour was observed (Jambon, 2006). Recording user behaviour is a delicate process. Video and audio recording must be as unobtrusive as possible in order not to influence the behaviour of the subjects while, on the other hand, the consent of the users for their recording should be always obtained. In addition, questions related to the frame of the recorded scene, viewing angle, and movement of the camera are significant. It must be stressed that there is a trade off between capturing the interaction with a specific device and capturing the overall scene of the activity. For example, often, crucial details may be missing from a video if recording the scene from a distance. Therefore, this video has to be complemented by other sources of related information, like screen captures of the devices used.

In order to conduct a successful usability evaluation, apart from collecting activity data, techniques and tools are needed for analysis of the collected information. In the last years, new usability evaluation techniques have emerged, suitable for mobile applications. Many of these methods focus mainly on user interaction with the mobile device, missing interaction between users, and user interaction with the surrounding environment.

Taking into consideration these aspects, the aim of this chapter is to discuss techniques and tools used first, for collecting data during usability evaluation studies of mobile devices, and then for the analysis of these data. In the process, a combination of a screen capturing technique and some tools that can be used for analysis of data of usability studies are presented.

BACKGROUND

The usability of a product has been traditionally related with the ease of use and learn to use, as well as with supporting users during their interaction with the product (Dix, Finley, Abowd, & Beale, 2003; Schneiderman & Plaisant, 2004). There have been many attempts to decompose further the term and render it operational through attributes and apt metrics. According to ISO 9241-11 standard, usability is defined as the "*extend to* *which a product can be used with effectiveness, efficiency and satisfaction in a specified context of use*" (ISO 9241). According to this view, a product's usability is directly related to the *user*, the *task*, and the *environment*. Consequently, usability cannot be studied without taking into consideration the goals and the characteristics of typical users, the tasks that can be accomplished by using the product, and the context in which it is going to be used. Making a step further on defining usability, the same standard suggests three potential ways in which the usability of a software product can be measured:

a. By analysis of the *features of the product* required for a particular context of use. Since ISO 9241 gives only partial guidance on the analysis process, in a specific problem there can be many potential design solutions, some more usable than others.
b. By analysis of the *process of interaction*. Usability can be measured by modeling the interaction with a product for typical tasks. However, current analytic approaches do not produce accurate estimates of usability since interaction is a dynamic process which is directly related to human behaviour that cannot be accurately predicted.
c. By analyzing the *effectiveness and efficiency*, which results from use of the product in a particular context, that is, measuring performance as well as the satisfaction of the users regarding the product.

Having in mind the three perspectives, there is a need for combining methods that capture the specific situation of use in a specific domain. Usability evaluation methods can be grouped in four categories (Nielsen, 1993): *Inspection*, *user testing, exploratory, and analytic methods.* Many techniques have been devised along these lines and have been extensively used in usability evaluation of desktop applications. Therefore the first approach in evaluating mobile applications was to apply these existing techniques. Such an approach can be found in Zhang and Adipat's (2005) survey of usability attributes in mobile applications which identified nine attributes that are most often evaluated: learnability, efficiency, memorability, user errors, user satisfaction, effectiveness, simplicity, comprehensibility, and learning performance. Such an approach is, however, limited, given the special characteristics of mobile devices with respect to desktop environments (Kjeldskov & Graham, 2003).

The mobile applications introduce new aspects to evaluate. The evaluation cannot be limited only to the device (typical scenario in desktop applications) but it must be extended to include aspects of context. The context in which the application is used is highly relevant to usability issues and often bears dynamic and complex characteristics. There is the possibility that a single device is used in more than a single context, in different situations, serving different goals and tasks of a single or a group of users. Also, group interaction, a common characteristic in mobile settings, gives a more dynamic character to the interaction flow of a system and increases the complexity of the required analysis as well as the necessity of observational data.

Along these lines, a new breed of methods for usability evaluation has been proposed (Hagen, Robertson, Kan, & Sadler, 2005; Kjeldskov & Graham, 2003; Kjeldskov & Stage, 2004). The process of selecting appropriate usability attributes to evaluate a mobile application depends on the nature of the mobile application and the objective of the study. A variety of specific measures (e.g., task execution time, speed, number of button clicks, group interactions, seeking support, etc.) have been proposed to be used for evaluation of different usability attributes of specific mobile applications. In the next section problems of data collection during mobile usability studies will be discussed.

Data Collection Techniques

A significant step during a usability evaluation study is to collect appropriate observational data to be analyzed. Hagen, Robertson, Kan, and Sadler (2005) classify the data collection techniques for mobile human-computer interaction in three categories: (a) *Mediated data collection (MDC)*, access to data through participant and technology, *do it*—the user makes himself the data collection; *use it*—data is collected automatically through logs; *wear it*—user wears recording devices that collect the data. (b) *Simulations and enactments (SE)* where some form of pretending of actual use is involved and (c) *combinations* of the above techniques. A review of different techniques of data collection, according to Hagen, Robertson, & Kan (2005) is shown in Table 1.

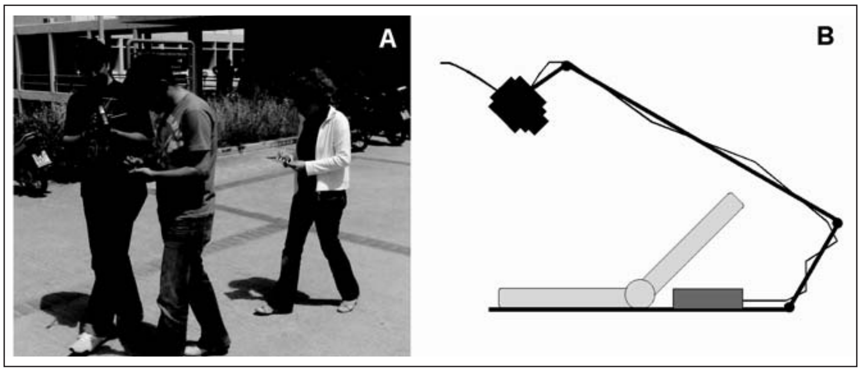
The data that are collected by these techniques come either directly from the user (through interviews, questionnaires, focus groups, diaries, etc.), by the evaluator (i.e., notes gathered during the experiment, observation of videos, etc.) or by raw data (log files, etc). All types of data need to be analyzed in order to become meaningful. Such data, in most cases, are in the following forms:

- *Log files* which contain click streams of user actions. These data can be derived by the application itself or by an external tool that hooks into the operating system message handler list. The latter case for mobile devices requires many system resources and therefore is not technologically feasible today, even in the most powerful mobile devices, like PDAs.
- *Audio/video recordings* of the users made through various means, like wearable mini cameras and/or audio recorders, static video cameras, operator or remote controlled cameras, from close or a far distance.
- *Screen recordings* by video cameras or by direct screen capturing through software (running on the device) the interaction flow in form of screen snapshots. This is a sequence of image representations of the user interface at certain instances that are usually taken at varying frequencies, usually a few snapshots per second. The screen snapshots can be stored either locally on the device (since it is feasible to store a large amount of data in memory cards) or on a central server over a wireless network connection

Table 1. Existing techniques for data collection used in studies of mobile technology. Adapted from Hagen, Robertson, and Kan (2005). F=Field, L=Laboratory, MDC=Mediated Data Collection, SE=Simulation and Enactments

Method	Description	Site*	Category
Artefacts (e.g. documents)	The use of objects or documents as sources for data collection. They may be objects (or photos of objects) from daily life or documents that users have created with devices being tested.	F	MDC
User Diaries	Users document information about their actions or thoughts, or impressions, often daily, for a period of time. Entries can be open and interpretive, or highly structured depending on the study.	F	MDC
Emulators	Emulators on desktop computers are used to simulate the interface of a potential mobile application.	L	SE
Focus Groups	Small groups of people are facilitated in unstructured discussion about an issue.	F, L	SE
Heuristics	Heuristics, often usability guidelines or design principles are applied by expert users to predict usability problems.	F, L	SE
Interviews	Interviews capture subject data from talking directly to participants. They can be open or structured and conducted in the field (including contextual interviews), online, over the phone and in labs.	F, L	SE
Log File analysis	Use logs are generated automatically (such as internet log files) or from systems specifically developed to capture content data and meta data.	F	MDC
NASA Task Load Analysis	Used in usability testing to determine work load.	F, L	SE
Observation/ Shadowing	Observation is used in field studies to capture use in context and can include, covert observation, participant observation, observing a place, or following a person. Data collection can include note taking, photography, and video.	F	MDC
Online data	Researchers gain access to information about the lives of users, and use practices from websites, forums and mailing lists.	F	MDC
Questionnaires	Quantitative or qualitative questionnaires are used to collect user opinions, feedback in evaluation, create user profiles or collect data about existing use practices. They can be done in person, or via phone or web.	F	SE
Role playing	Users and researchers play out different roles, or act out tasks or scenarios to explore existing and future use concepts	F, L	SE
Scenarios	Scenarios provide information about use situations giving examples of how technologies are used in practice.	F, L	SE
Think-Aloud	Participants describe out loud what they are thinking while they complete tasks using a device or prototype.	F, L	SE

Figure 1. A) Shadowing technique (see also Kjeldskov & Stage, 2004); B) Recording screen with wireless camera (see also Betiol & de Abreu Cybis, 2005)



Screen recordings of mobile devices are invaluable resources that can greatly help evaluators identify usability problems. Various techniques can be used for capturing the screen of a mobile device: One is the recording of the screen by using a mini wireless camera (Figure 1B). It can be very helpful in cases of individual users but it is not suitable in the case of an application that involves beaming actions (e.g., Bluetooth, infrared) and/or interaction with the physical space because it can influence negatively the use of the device and can create obstacles in the infrared beams, sensors, or readers attached to the device (i.e., to an RFID reader). The main advantage of this technique is that the camera records, besides the screen, the movements of the users fingers or stylus, capturing valuable data identifying potential interaction problems (for example, the user hesitates to click something because the interface or the dialogs are confusing).

An alternative technique is the shadowing technique which can effectively work for individual users (Figure 1A). Again, this technique is not suitable for group activities, where the subjects often form groups and move continuously. Even in cases that it is considered possible to record properly, there could be many events missing because of the frequent movements of the subjects or the shielding of the screen by their body and hands.

The direct observation technique has also certain limitations (Cabrera et al., 2005; Stoica et al., 2005) because the observer must distribute his attention to many subjects. In case there are observers available for each user they will restrict the mobility of the users and they will distract their attention when being in so close range. Consequently, all these techniques impose the presence of the observer to the users, thus affecting their behavior.

Another significant issue that directly affects the usability evaluation is related to the location in which the study is conducted. There are many arguments in favour of *field usability studies* (Nardi, 1996; Kjeldskov, Skov, Als, & Hoegh, 2004; Zhang & Adipat, 2005; Kaikkonen, Kallio, Kekalaien, Kankainen, & Cankar, 2005). Comparative studies between laboratory and field evaluation studies have drawn, however, contradictory conclusions. In a recent survey of evaluation studies of mobile technology (Kjeldskov & Graham, 2003), 71% of the studies were performed in the laboratory, which revealed a tendency towards building systems based on trial and error and evaluating systems in controlled environments at the expense of studying real use of them. So the question of what is useful and what is perceived problematic from a user perspective often is not adequately addressed.

In summary, in order to conduct a usability evaluation of a mobile application/system, there is a need to take into consideration the attributes that are going to be measured, the data collected for these measurements, the location in which the evaluation will take place, and finally, the appropriate tools to analyze them, having always in mind the user and the context of interaction.

Data Analysis

Usability evaluation of mobile applications is more complex than desktop software evaluation since new characteristics such as group activity and the interaction with the surrounding environment need to be taken into consideration. In order to acquire an understanding of group activity and performance, huge amounts of structured and unstructured data of the forms discussed in the previous section need to be collected. These data should capture the activity of subjects, including their movements, facial expressions, gestures, dialogues, interaction with the devices, and objects in the environment. Analysis of these data require special attention on details as well as the context of use, thus it can be a tedious process which can be facilitated by a suitable analysis tool (Benford et al., 2005).

Various tools have been developed to support usability evaluation studies and, in general, to record and annotate human activity. These tools often handle video and audio recordings and synchronize them with text files, containing hand-taken notes. This combination creates a dataset that is rich in information which is then annotated through an adequate annotation scheme, which creates quantitative and qualitative measures of the observed user-device interaction. Typical examples of such tools are: the *Observer XT* (Noldus, 2006), *HyperResearch* (Hesse-Biber, Dupuis, & Kinder, 1991; ResearchWare, 2006), *Transana* (Transana, 2006), *NVivo* (QSR, 2006; Rich & Patashnick, 2002; Welsh, 2002), and *Replayer* (Tennent & Chalmers, 2005). From them, only *Replayer* and *Observer XT* have special provisions for mobile settings. The extra characteristics in evaluation of mobile applications (group activity and interaction with the surrounding space) demand the extended use of multimedia files that thoroughly capture the activity. Thus, there is a need for a tool that combines and interrelates all of the observational data in a compact dataset and gives to the usability expert the ability to easily navigate them from multiple points of view (access in user—device interaction, access in user—space interaction).

All of the tools utilize video sources at a different extend, with the exception of *NVivo* that focuses more in textual sources. *NVivo* allows linking of evaluator's notes with video extracts, without permitting more fine grained handling of video content. On the other hand, *HyperResearch* and *Transana* do support flexible handling of video sources but they do not allow the integration and synchronous presentation of multiple video sources in the same study. Thus, *NVivo*, *HyperResearch*, and *Transana* cannot successfully respond to the extra characteristics of mobile applications. On the other hand, *Replayer* is a distributed, cross platform toolkit that allows the integration of multiple video sources and presents analysis data in various forms such as histograms and time series graphs. Although *Replayer* efficiently supports usability analysis of mobile applications, its failure to handle and to compare data that come from various studies makes it not suitable for cases of multiple studies in which there is need to aggregate and generalise the findings. On the contrary, *Observer XT* is a powerful commercial tool, widely used in observation studies, that enables the synchronous presentation of multiple video files and also the derivation of overall results about the activity of multiple subjects. Although *Observer XT* meets the requirements of new characteristics of mobile applications, its use requires a prior lengthy training period.

A tool that has been especially adapted for analysis of data from mobile applications' evaluation studies is the *ActivityLens* which attempts to tackle some of the limitations of existing tools. Its main advantage is its ability to integrate multiple heterogeneous qualitative but also quantitative data. It allows the usability expert to directly access the collected data, thus to simultaneously focus on users' movements on the surrounding environment and user-device interaction. To sum up, *ActivityLens* supports analysis of collected data and produces results that cover the overall activity concerning all the participants.

Weitzman and Miles (as cited in Berkowitz, 1997) suggest that a criterion for the selection of

Table 2. Characteristics of usability evaluation tools

	Multiple multimedia sources	Aggregated results from multiple studies
Observer XT	☑	☑
HyperResearch		☑
Transana		
NVivo		☑
Replayer	☑	
ActivityLens	☑	☑

an adequate analysis tool is related to the amount, types, and sources of data to be analyzed and the types of analyses that will be performed. In Table 2 a description is provided about how the tools support the extra characteristics of usability evaluation studies of mobile applications.

Data Analysis through ActivityLens

ActivityLens is a tool that embodies features especially designed for usability evaluation of mobile applications. *ActivityLens* is an evolution of the earlier Collaboration Analysis Tool (ColAT) (Avouris, Komis, Margaritis, & Fiotakis, 2004; Avouris, Komis, Fiotakis, Margaritis, & Voyiatzaki, 2005), originally designed for video analysis of collaborative learning activities. It was found particularly suitable for the proposed approach which involves multiple perspectives of the activity, based on different multimedia data.

In *ActivityLens*, all the collected data are organized into *Studies*. An example of a *Study* is the usability evaluation that was conducted in a Historical Cultural museum, described in the next sections. The tool allows *Projects* that belong to a specific *Study* to be defined. A *Project* is defined by the evaluator and can have different perspectives depending on the situation. For example, a *Project* can be defined as the set of data gathered from various groups over a set period of time, or it can be defined as a set of data of a specific group of users.

These data can be video and audio files, log files, images, and text files, including hand-taken notes of the observers. *ActivityLens* supports almost all the common video and audio file formats including file types that are produced by mobile devices such as .mp4 and .3gp. The observed activity is reported in an XML log file. This file describes the activity as a set of events, reported in sequential order, following this typical structure:

<event id>, <time-stamp>, <actor>, <tool>, <event-description>, <type of event>, <comments of evaluator>

The log file events are presented via a simple spreadsheet view in order to be easily accessible for inspection and annotation. In addition, *ActivityLens* permits integration and synchronization of the collected multimedia files.

All the data can be reproduced and annotated on-the-fly in order to highlight interesting events. An example is shown in Figure 2, in which an overview video and a PDA screen are synchronized and annotated. The annotation of the observed events is based on a classification scheme defined by the evaluator. For example, an evaluator is analyzing videos that describe the activity of a group of students that try to solve a problem. During the activity some students propose ways to solve the problem and argue about it. Thus, one representative type of event could be defined as "Proposal." For usability studies, an evaluator can define typologies based on usability attributes, concerning for instance, user errors, comments expressing subjective view, and events marking successful completion of tasks.

Figure 2. The usability evaluation tool—ActivityLens

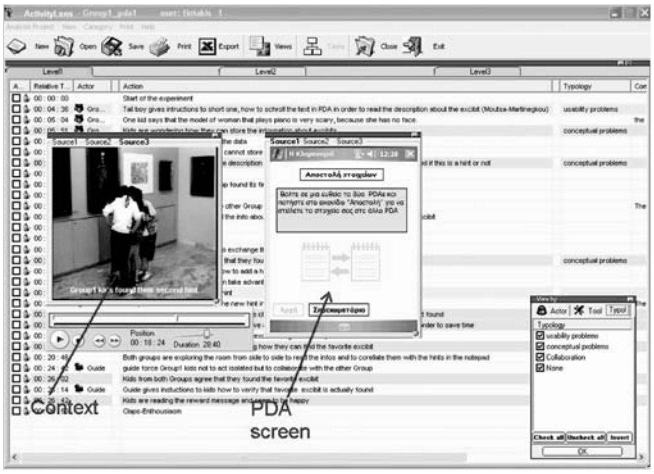
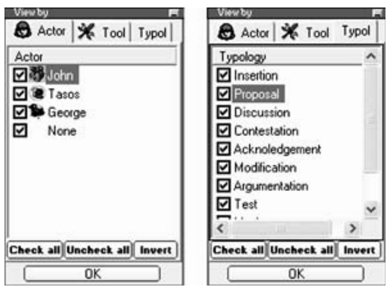


Figure 3. Event filtering tool through ActivityLens

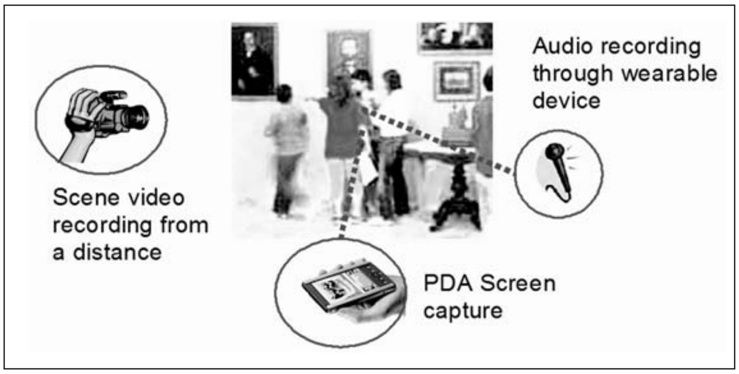


ActivityLens provides the evaluator with the ability to reduce the huge amount of collected data through an event filtering mechanism. This feature is of high importance because it helps the evaluator to focus on interesting sequences of events and makes them emerge from the "noise." The evaluator is allowed to define criteria for specific Actors, tools used, and types of events or any combination between them. For example, the evaluator can choose to view all occurrences of "Proposals" made by "George" or "John." The criteria selection tool is shown in Figure 3.

PROPOSED METHODOLOGY

Based on the outlined data and analysis requirements, in this section a methodology suitable for usability evaluation of mobile applications is proposed. This method is proposed for applications deployed in places like museums, libraries, and so forth, in which groups of users interact among themselves and with the environment, in various ways. These semi-public spaces represent 'living organisms' that project, in a visible and tangible form, the various facets of information. For example, in a museum such applications assist the visitors in discovering and acquiring knowledge. A museum can be characterized as an ecology (Gay & Hebrooke, 2004) that is constituted by two main entities, the exhibits and the visitors, populating the same space. Items of the collection are exhibited to visitors, who react by discovering them in a way that is, at a large extent, influenced by the surrounding space. Also, visitors usually interact with each other, for example, because they com-

Figure 4. Sources of observational data



ment the exhibits independently from the use of technology. This methodology involves, initially, the preparing study phase, the recording activity phase, and then the analysis of the activity.

Preparing the Study

Usually, activities that are expected to take place in semi-public spaces are desirable to be conducted in the field. For example, visitors inside a museum enjoy an experience that cannot be fully reproduced inside a laboratory. Therefore, the evaluator needs to conduct a study in a representative place, which should be adapted accordingly without disturbing its normal operation. Issues to be tackled are related with technological restrictions (e.g., wireless network infrastructure), recruitment of an adequate number of typical users, the extent of the study, and so forth. Consequently, it is evident that the preparation phase of the evaluation is a very important one, as it builds the foundation for a subsequently successful study.

Recording Activity Phase

A prerequisite in such environments is the low level of activity interference by the observers in order to minimize the behavioral change caused to the participants by the uncomfortable feeling of being observed and thus "disorienting the balance" in the ecology. The proposed recording activity includes an innovative combination of existing data gathering techniques in order to achieve the considered goal. The sources of data (Figure 4) include: (a) screen recordings of the mobile devices, (b) audio recordings using wearable recorders, (c) video recordings from the distance, where the camera is operated by an operator or preferably by remote control, and as complementary source (d) interviews and questionnaires to the users. A brief discussion of the process of collecting these data is included next.

(a) Screen Grabbing on the Mobile Device

In order to tackle problems related to the application nature (collaboration, interaction with the environment) it is proposed that the mobile device also be used as a screen recording device. The collected information can be in the form of screen-shots or aggregated in a low frame-rate video. The main requirement for a mobile device to become a screen recording device is that it must run a multitasking operating system in order to allow a background process to run in parallel with the main application. At the current technological status, this is the case for most mobile devices (PDAs and smart phones), as the main operating systems are multitasking: Symbian OS, Windows Mobile, Palm OS (version 6.0 onwards), Java OS, and so forth. Also, the needs of the market drove the mobile devices to handle large amounts of data that have to be consulted, edited, and updated by the user while speaking, browsing, watching TV, and so forth. As a result, mobile devices evolved from a single process, sequential to multitasking, and obtained increased storage capacities which permit the users to store a lot of information on them. Therefore, a mobile device can capture, by a parallel process, the screen and either save the pictures on their memory or send them directly to a server via a wireless connection.

A prototype application that is suitable for the Pocket PC/Windows Mobile environment has been developed and runs in parallel with the application which has to be evaluated. It captures screen snapshots and stores them on the device at a predefined time interval. In the tests, a compressed quarter VGA (240x320) screen shot was at most 32 KB that at a rate of 4 per second lead to a needed storage of about 450 MB/hour. It must be stressed that far better compression rates can be achieved by using video encoders.

The decision to grab the screen with a steady frequency and not per number of events, that would make sense in order to stop recording when the device is not used, was imposed by the technical current limitation: the scarce support for global system hooks on the Windows Mobile operating system. The lack of support is due to the fact that such hooks can critically affect the performance of the device.

(b) Audio Recording with Wearable Devices

Audio can capture dialogs between users that express difficulty in interacting with the application and the environment, or disagreement. Audio recordings can often reveal problems that users do not report during interviews or questionnaires.

The audio recordings from the inbuilt microphone of the video camera are sometimes not very useful due to the noise and to the fact that usually the dialogues are in a low voice. Also, the distance between the subject and the camera does not allow recording of good quality sound. The ideal solution would be that the mobile device itself could record both the screen and the audio. Unfortunately, this is not feasible because of several reasons:

- The performance of the device degrades significantly by having two background processes running simultaneously, the one related to screen grabbing, discussed in (a) and the one to audio recording.
- The sounds that are produced by the device itself, in most cases, cover any other sound in the surrounding environment (i.e., a narration played back covers the dialogue.
- The storage might be a problem. Depending on the audio quality and compression used, 1 hour of recorded sound can take from 50 MB to 700 MB.

For these reasons, it seems that the most suitable solution is to use a wearable audio recorder that can store several hours of sound. These devices are very light; they weight less than 50 grams, including the battery. The user can wear it with the help of a neck strap or put it in a pocket and adjust a clip microphone. The wearable audio recorders guarantee that rich information concerning the dialogs between the subjects will not be lost, collaborating and interacting with the application and the environment.

(c) Discrete/Unobtrusive Video Recording

To complement the dialogs and the screen recordings, it is necessary to capture, in video, the ensemble. From this video, recording the context of the events, the social interactions between the group members (peers) and/or between groups can be depicted. In order to decrease as much as possible the level of obtrusion, the camera must be preferably maneuvered through remote control (allowing zoom and angle changes) or at least by a cameraman that will keep a large enough distance from the activity in order not to disturb the users. Often, many video recordings may need to be made from various angles, distance, or focusing in different aspects. These may be mixed in a single video stream if adequate equipment is used, or, more often, may be kept as separate sources of information. By studying these video recordings the evaluator can obtain a clear idea about the place in which the activity took place.

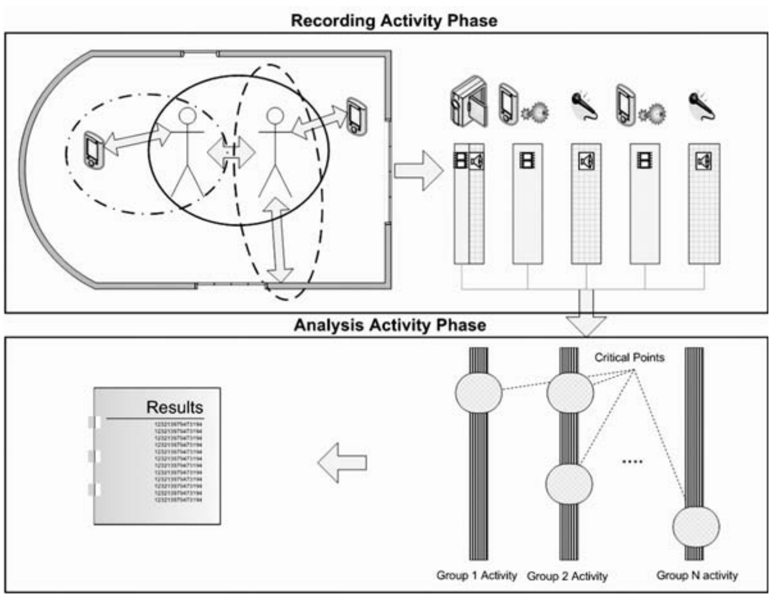
(d) Interviews and Questionnaires

Considering that the sources constitute the objective information, the users' subjective view through interviews and questionnaires also need to be obtained. Through these sources, which vary depending on the situation, someone can formulate results regarding *user's satisfaction*, *learning performance*, and so forth; attributes sometimes difficult to obtain simply through observation.

Analysing Activity Phase

The purpose of the analysis is to identify instances of use of the devices and the infrastructure, which identify usability problems of the technology used. Analysis of recorded activity of groups in semi-public spaces is not a simple process. Researchers have not only to focus just on the devices but to take into account more complicated issues concerning the interaction between groups, the interaction between peers in a certain group, and the interaction with the surrounding space. This analysis has to be meticulously performed in order to cover the above issues. During analysis all the collected sources that describe the group activity have to be combined and iteratively inspected. Initially, a quick inspection of recorded activity helps usability experts to isolate the segments that need thorough analysis. Then, detailed inspection of these segments is required to interpret the observed interaction and depict the usability problems. This process can help usability experts to detect certain critical points of interaction that can be further examined in order to measure their frequency and dispersion between groups and to be clear how they affect the use of mobile applications. The proposed methodology concerning the recording and analysis process can be seen in Figure 5.

Figure 5. Recording and analysis phases of proposed methodology: Interesting incidents are observed in the media files and are cross-checked for better understanding. These incidents are analyzed in terms of device and activity usability issues.



EVALUATING USABILITY OF A COLLABORATIVE CONTEXT AWARE EDUCATIONAL GAME

An example of a study in which the proposed technique was applied was a usability evaluation of a collaborative mobile learning application supported by PDAs in a cultural-historical museum (Tselios et al., 2006). The study involved 17 students of the 5th-grade of an elementary school (11 years old) who were invited to visit the museum and use the prototype of an educational application that was temporarily installed there. All the students were familiar with the use of mobile phones but they had no former experience with PDAs. Furthermore, most of them described themselves in a pre-study questionnaire, as users of desktop computer systems on a daily basis.

The study took place in two of the museums' halls in which portraits and personal objects of important people of the local community were exhibited. First, a short introduction to the activity was provided by a member of the research team who undertook the role of the guide. The educational activity was designed in a way that students were motivated to read information about these important people and collaboratively search in order to locate a specific exhibit according to the activity scenario. The children were divided in two groups and each group consisted of two teams of 4 or 5 children each. Each group participated in a different session for approximately 1 hour.

In order to achieve the scenario's goal, each team was provided with a PDA equipped with a RFID tag reader. They used this equipment to locate hints that were hidden inside textual descriptions of the exhibits. These were obtained by scanning the exhibit RFID tags. The students could store the hints it in a notepad of the PDA. After collecting all or most of the hints the teams were encouraged to share their hints, through beaming, to each other.

Then the students, using the found information, had to locate a specific-favorite exhibit which matched the description provided by the hints. When two teams agreed that they had found the favorite exhibit, they checked the correctness of their choice by scanning with both PDA's the RFID tag. A correct choice was indicated by the system with a verification message while a wrong one suggested a new search. When the study was over each student was requested to answer a set of questions related to the group activity in the museum.

Preparation of the Evaluation Study

During the preparation of the study the museum was contacted and the permission to run the evaluation study was obtained. The space of the museum was examined well in advance (e.g., for determining wireless network setup options) and afterwards a small scale pilot was run in a simulated environment in order to check the suitability of the technological infrastructure. In order to ensure the participation of subjects, a school in the vicinity of the museum was contacted and participation of a school party was requested for the study.

Collecting Data

In order not to miss important contextual information, three video cameras were used in this study. Two of them were steadily placed in positions overlooking the halls while the third one was handled by an operator who tenderly followed the students from a convenient distance. One student per team wore a small audio recorder in order to capture the dialogues between them, while interacting with the application and the environment. Furthermore, snapshots of the PDA screens were captured during the collaborative activity and stored in the PDA's memory. After the completion of the study, the guide, who was a member of the evaluator team, had an interview with the students, asking them to provide their opinion and experiences from the activity in the museum; while back at school a week later, their teacher asked them to write an essay describing their experience.

Analysing Data with ActivityLens

In order to analyse all the collected data according to the proposed methodology, ActivityLens, that has already been effectively used in similar studies was used (Cabrera et al., 2005; Stoica et al., 2005).

The main reasons that ActivityLens was used among the discussed tools was its capacity of organizing observations into Studies (collection of projects) and its ability to present multiple perspectives of the whole activity (by integrating multiple media sources). Although Observer XT provides even more capabilities than ActivityLens, the choice of ActivityLens seemed to fit better the specific use case since its use did not require a long training time. In addition, ActivityLens permits easy access to the activities of the subjects recorded in different data sources.

Three usability experts, with different levels of experience, analysed the collected data in order to increase the reliability of the findings. Initially, a new *ActivityLens Study* including four projects (each project concerns the observations of a team) was created. The integrated multimedia files were extensively studied and the most interesting situations were annotated. It must be clarified that it was not wanted for the behaviour of each individual team member to be studied but wished that the performance of the whole team be evaluated. The performed analysis through ActivityLens revealed several problems related to the children's interaction with the device and the overall setting, given the surrounding physical space and groups.

Several problems were identified when the students interacted with the handheld devices.

The analysis indicated that almost all the groups could not successfully scan the Exhibits tags in their initial attempts and get information about the exhibits. The RFID tags were located underneath each exhibits label. Since the users had no clear indication of where to place the tag scanner, some of them experienced difficulties interacting with them. Also, there was an unexpected delay in the scanning process between tag and PDA (the PDA needed about 2 seconds to scan the tag). While from the scene, video recording, it seemed that the user was repeatedly scanning the same label, combining this with the PDA screen recording gave the real reason of this behavior—repeated unsuccessful tries to scan the tag. The users learned after a few frustrating attempts that they should target the center of the tags and hold the device for a couple of seconds.

A problem that troubled a specific group was the use of a scrollbar in the textual description of the exhibits. The users were not familiar with the procedure of scrolling on a PDA and they repeatedly discussed it amongst themselves. This problem was identified through the combined use of the audio and screen recording and was not visible from the scene video.

An unexpected problem was related to the content of some exhibits descriptions. They

Figure 6. A) Instance of user—RFID tag interaction problem. B) and C) Photos from the collaboration activity inside the museum

contained the word "hints" which confused the children and they were not sure if this was or was not a hint that they could add to the notepad. This was spotted from the complementary use of the overview video with the dialogue audio recordings. The problem was overcome by asking the help of the guide.

With the use of ActivityLens many problems that were related to the interaction with the physical space were managed to be detected. The most important one was that some of the exhibits tags were placed on the walls in such positions that they were not accessible by short students. In Figure 6 an instance of this problem is shown.

Another interesting element that was made clear through the students' dialogues and the videos was that in a certain area of the room an exhibit inspired fear to some of the children (e.g., a faceless piano player). Particularly, one student was clearly afraid to get near the puppet and said to the other members: "I am not going near her. She is very scary!!! Look at her, she has no face!" This situation made the team avoid that area, which contained exhibits with useful information for the activity.

The children that participated in the study often expressed their concern about being delayed in their play due to the presence of other museum visitors (at a certain point an independent school party crowded the hall). Through the audio it was obvious that the kids expressed their frustration because they were delayed in playing the game and the visitors, because they were disturbed by the kids. These problems escape from the traditional usability analysis that focuses only on the device, because they contain the interaction between the user and the surrounding physical space.

The third dimension of the evaluation concerned investigation of the collaborative nature of the activity and the learning performance. An interesting observation was that by having two teams searching for hints at the same time, and the fact that one of the teams was more successful than the other, constituted a powerful motivation for the second team to search for hints. This was observed from the complementary scene video (pinpointing the event) and the dialog recordings (exclamations, etc.). Also, that some kids were too excited in using the PDAs and did not allow anyone else to use them was observed. Thus, disputes over use of the device influenced negatively the team spirit. From the audio streams, the disappointment of the kids that were not allowed to use the device were managed to be spotted.

Regarding the learning performance through the audio files and the PDA's screen it was found that one team was not reading the descriptions to locate the hints but they were searching for the parentheses that indicated the existence of a hint. It must be said that the solution with the parentheses and not colored text was adopted because it was wanted that those specific situations be avoided, but this did not actually work in all teams. In the future version, the hints will be visible only when the users click on them inside the description of the exhibits.

The results are also based on a study of questionnaires, independently of the ActivityLens analysis. In this point, the limitation of ActivityLens in analyzing user questionnaires has to be underlined. This weakness is a matter of further development and research.

In order to have a general view about the educational value of the activity when the children returned to their classrooms, they wrote an essay in which they reported on the museum experience. The teacher's view after going through these texts was that almost all the kids that participated in the activity learned something meaningful in a funny and enjoyable way. However, a more systematic study on these issues should involve a more quantitative experimental approach through a pre and post-test questionnaire and a control group.

CONCLUSION

This chapter has presented a brief overview of usability evaluation techniques for mobile applications, including collection of multiple observational data and their analysis. Due to the growing use of mobile devices, it is evident that there is a need for established techniques that support the collection and analysis of data while conducting usability evaluations. Since there are considerable differences between desktop and mobile environments, researchers are obliged to develop and fine tune these new techniques. Through this chapter a methodology for evaluating mobile applications focusing on collection and use of observational data was proposed. The proposed methodology was demonstrated through a usability study of an educational game in a Historical Museum.

The proposed recording activity technique can be characterized as unobtrusive regarding the users and allows evaluators to study the activity in conditions as close as possible to the typical conditions of use of the application, through various perspectives. The ActivityLens tool was used for analysis of the collected data which facilitates interrelation and synchronization of various data sources and was found particularly useful since the collected data were of particularly high volume and often a finding was based on a combination of data sources. The methodology revealed usability problems of the application as well as issues about collaboration and interaction with the environment that would not be easy to discover in the laboratory and without the combined use of the multiple media data.

Studies that take place in semi-public spaces and involve groups of people have to tackle various problems. In most cases the willingness of people but also the availability of spaces is difficult to be guarantied for the long periods of time. Researchers that conduct such studies have to be as unobtrusive as possible to the users and pay special attention in order to minimize interference with the environment.

A limitation of the proposed approach is that it requires the users to carry light equipment (audio recorders) and also that a screen capturing software had to be installed in the mobile devices. However, these limitations did not inhibit the users to act naturally and recreate a realistic but controlled context of use. The typical studies of the proposed approach lasted a short time and thus, it is difficult to measure long term usability aspects like memorability and long term learning attitudes. It is still under investigation how to extend this technique to long term mobile usability studies involving different contexts of use.

What is however, missing from the story is an analysis scheme that can describe user interaction with the surrounding physical and information space and metrics that map usability attributes. Such a scheme would describe usability as a set of attributes that refer to interaction with the device, interaction with the space, and group interactions. This scheme could be supported by a tool like ActivityLens which facilitates easy navigation of the collected media data, allowing creation of pointers to incidents in the data, justifying the calculated values of the usability attributes. Definition of such a scheme should however, be the result of a wider research community process.

REFERENCES

Aittola, M., Parhi, P., Vieruaho, M., & Ojala, T. (2004). Comparison of mobile and fixed use of SmartLibrary. In S. Brewster & M. Dunlop (Eds.), *Proceedings of 6th International Symposium on Mobile Human-Computer Interaction (Mobile HCI 2004)* (pp 383-387). Berlin: Springer.

Aittola, M., Ryhänen, T., & Ojala, T. (2003). SmartLibrary—Location-aware mobile library service. In L. Chittaro (Ed.), *5th International Symposium on Human-Computer Interaction with Mobile Devices and Services (Mobile HCI 2003)* (pp. 411-416). Berlin: Springer.

Avouris, N., Komis, V., Fiotakis, G., Margaritis, M., & Voyiatzaki, E. (2005). Logging of fingertip actions is not enough for analysis of learning activities. In *Proceedings of AIEDs Workshop on Usage Analysis in learning systems*. Retrieved February 27, 2007, from http://lium-dpuls.iut-laval.univ-lemans.fr/aied-ws/.

Avouris, N., Komis, V., Margaritis, M., & Fiotakis, G. (2004). An environment for studying collaborative learning activities. *Journal of International Forum of Educational Technology & Society, 7*(2), 34-41.

Benford, S., Rowland, D., Flintham, M., Drozd, A., Hull, R., Reid, J., Morrison, J., & Facer, K. (2005). Life on the edge: supporting collaboration in location-based experiences. *Proceedings of the SIGCHI conference on Human Factors in computing systems CHI 2005* (pp. 721-730). New York: ACM Press.

Berkowitz, S. (1997). Analyzing qualitative data. In J. Frechtling & L. Sharp Westat (Eds.), *User-friendly handbook for mixed method evaluations.* Retrieved February 27, 2007, from http://www.ehr.nsf.gov/EHR/REC/pubs/NSF97-153/start.htm

Betiol, H. A., & de Abreu Cybis, W. (2005). Usability testing of mobile devices: A comparison of three approaches. In M. F. Costabile & F. Paterno (Eds.), *Proceedings of IFIP TC13 International Conference on Human-Computer Interaction (INTERACT 2005)* (pp. 470-481). Berlin: Springer.

Cabrera, J. S., Frutos, H. M., Stoica, A. G., Avouris, N., Dimitriadis, Y., Fiotakis, G., & Demeti, K. (2005). Mystery in the museum: Collaborative learning activities using handheld devices. In M. Tscheligi, R. Bernhaupt, & K. Mihalic (Eds.), *Proceedings of the 7th International Conference on Human Computer Interaction with Mobile Devices & Services (Mobile HCI 2005)* (pp. 315-318). New York: ACM Press.

Dey, A. (2001). Understanding and using context. *Personal and Ubiquitous Computing Journal*, *5*(1), 4-7.

Dix, A., Finley, J., Abowd, G., & Beale, R. (2003). *Human-computer interaction* (3rd ed.). Hertfordshire: Prentice Hall.

Fouskas, K., Pateli, A., Spinellis, D., & Virola, H. (2002). *Applying contextual inquiry for capturing end-users behaviour requirements for mobile exhibition services.* Paper presented at the 1st International Conference on Mobile Business. Athens, Greece.

Gay, G., & Hebrooke, H. (2004). *Activity-centered design. An ecological approach to designing smart tools and usable systems.* Cambridge, Massachusetts: MIT Press.

Hagen, P., Robertson, T. & Kan, M. (2005). *Methods for understanding use of mobile technologies.* Technical Report. Retrieved September 20, 2006, from http://research.it.uts.edu.au

Hagen, P., Robertson, T., Kan, M., & Sadler, K. (2005). Emerging research methods for understanding mobile technology use. In *Proceedings of the 19th Conference of the Computer-Human Interaction Special Interest Group (CHISIG) of Australia on Computer-human interaction: Citizens online: Considerations for today and the future OzCHI 2005* (pp. 1-10). New York: ACM Press.

Hesse-Biber, S., Dupuis, P., & Kinder, T. S. (1991). HyperRESEARCH, a computer program for the analysis of qualitative data with an emphasis on hypothesis testing and multimedia analysis. *Qualitative Sociology, 14*, 289-306.

Jambon, F. (2006). Reality testing of mobile devices: How to ensure analysis validity? In *Proceedings of CHI 2006 Workshop on Reality Testing: HCI Challenges in Non-Traditional Environments.* Retrieved February 27, 2007, from http://www.cs.indiana.edu/surg/CHI2006/WorkshopSchedule.html

Kaikkonen, A., Kallio, T., Kekalaien, A., Kankainen, A. & Cankar M. (2005). Usability testing of mobile applications: A comparison between laboratory and field testing. *Journal of Usability Studies*, *1*(1), 4-16.

Kjeldskov, J., & Graham, C. (2003). A review of Mobile HCI research methods. In L. Chittaro (Ed.), *5th International Symposium on Human-Computer Interaction with Mobile Devices and Services (Mobile HCI 2003)* (pp. 317-335). Berlin: Springer.

Kjeldskov, J., & Stage, J. (2004). New techniques for usability evaluation of mobile systems. *International Journal of Human-Computer Studies, 60,* 599-620.

Kjeldskov, J., Skov, M. B., Als, B. S., & Hoegh, R. T. (2004). Is it worth the hassle? Exploring the added value of evaluating the usability of context-aware mobile systems in the field. In S. Brewster & M. Dunlop (Eds), *6th International Symposium* *on Mobile Human-Computer Interaction (Mobile HCI 2004)* (pp 61-73). Berlin: Springer.

Nardi, B. (1996). Studying context: a comparison of activity theory, situated action models, and distributed cognition. In B. Nardi (Ed.), *Context and consciousness: Activity theory and human-computer interaction* (pp. 69-102). Cambridge, Massachusetts: MIT Press.

Nielsen, J. (1993). *Usability engineering.* London: Academic Press.

Raptis, D., Tselios, N., & Avouris, N. (2005). Context-based design of mobile applications for museums: a survey of existing practices. In M. Tscheligi, R. Bernhaupt, & K. Mihalic (Eds.), *Proceedings of the 7th international Conference on Human Computer interaction with Mobile Devices & Services (Mobile HCI 2005)* (pp. 153-160). New York: ACM Press.

Rich, M., & Patashnick, J. (2002). Narrative research with audiovisual data: Video intervention/prevention assessment (VIA) and NVivo. *Int. Journal of Social Research Methodology 5*(3), 245-261.

Schneiderman, B., & Plaisant, K. (2004). *Designing the user interface: Strategies for effective human-computer interaction* (4th ed.). Boston: Addison Wesley.

Stoica, A., Fiotakis, G., Cabrera, J. S., Frutos, H. M., Avouris, N. & Dimitriadis, Y. (2005, November). *Usability evaluation of handheld devices: A case study for a museum application*. Paper presented at the 10th Panhellenic Conference on Informatics (PCI2005), Volos, Greece.

Tennent, P., & Chalmers, M. (2005). Recording and understanding mobile people and mobile technology. In *Proceedings of the1st International Conference on E-social science*. Retrieved February 27, 2007, from http://www.ncess.ac.uk/conference_05.htm/papers/

Tselios, N., Papadimitriou, I., Raptis, D., Yiannoutsou, N., Komis, V., & Avouris, N. (2006). *Design for mobile learning in museums.* To appear in J. Lumsden (Ed.), *Handbook of User interface design and evaluation for mobile technology.* Hershey, PA: IGI Global.

Welsh, E. (2002). Dealing with data: Using NVivo in the qualitative data analysis process. *Forum Qualitative Social Research Journal, 3*(2). Retrieved February 27, 2007, from http://www.qualitative-research.net/fqs-texte/2-02/2-02welsh-e.htm

Zhang, D., & Adipat, B. (2005). Challenges, methodologies, and issues in the usability testing of mobile applications. *International Journal of Human-Computer Interaction, 18*(3), 293-308.

KEY TERMS

ActivityLens: A usability analysis tool used to support usability studies for mobile and collaborative applications analyzing multiple media data.

Context: Any information that can be used to characterize the situation of an entity. An entity should be treated as anything relevant to the interaction between a user and an application, such as a person, a place, or an object, including the user and the application themselves. (Dey, 2001).

Context Aware: A device, a system, or an application that has the ability to sense aspects of context and change its behaviour accordingly.

Data analysis tool: A software package that supports extracting meaningful information and conclusions from collected data.

Data Collection: The process of gathering raw or primary specific data from a single source or from multiple sources.

Screen Recording: The operation of capturing the output of a devices' screen**.**

Semi-Public Space: A place which is public to people and imposes a set of common, and universally acceptable rules regarding their behaviour i.e. a museum, library, theatre.

Usability Evaluation: The process of assessing the usability of a given system or product.

Section V
Case Studies

This final section includes a selection of detailed case studies. These illustrate many of the concepts discussed in previous sections and cover the design of mobile technology for a closed environment (in this case, a hospital), the design of a memory-aid, an application for reviewing meeting records, and a tool for mobile collaborative reading, as well as the evaluation of learner satisfaction in a multiplatform learning system.

Chapter LX
UI Design for Mobile Technology in a Closed Environment

Kate Oakley
Carleton University, Canada

Gitte Lindgaard
Carleton University, Canada

Peter Kroeger
BRYTECH, Canada

John Miller
BRYTECH, Canada

Earl Bryenton
BRYTECH, Canada

Paul Hébert
Canadian Medical Association, Canada

ABSTRACT

This chapter reports on a case study linking several technology devices that monitor a range of vital signs in patients recently discharged to a hospital ward from the Intensive Care Unit (ICU). Apart from presenting an interesting technological challenge, this closed environment creates unique logistical and physical ergonomic challenges as well as cognitive and perceptual design problems for mobile technology. Devices include desktop computers, touch monitors, and several types of remote mobile devices including PDAs. A number of important design issues are addressed, such as deciding which *visual details can be safely eliminated from a small display, or if permission should be given to turn off the alarm functions, among others. Lack of direct access to users compromised the ecological validity of several parts of the evaluation and alternative evaluation methods had to be devised.*

INTRODUCTION

Designing applications for mobile devices is always an interesting challenge. Users demand more and more capabilities on devices continually decreasing in size, expect more and better integrated services, require the interaction model and display to resemble their desktop, and desire new applications to be usable without much investment in learning to use them. While present-day mobile computing is a far cry from Mark Weiser's (1991) conception of 'calm' computing (Rogers, 2006), mobile technologies do open the door to ubiquity and to fulfilling a variety of aims. For example, with a focus on seniors and vulnerable people, a number of technologies designed to track, sense, and alert have been introduced (Mynatt, Melenhorst, Fisk, & Rogers, 2004). However, deciding what to video, sense, and track is an ongoing issue. As Rogers (2006) so aptly asks "is it right to be videoing and sensing people when they are sleeping, eating, etc., especially when they are not at their best?" (see also, Anderson *&* Dourish, 2005). While the motivation behind such applications is innocently altruistic, it does not take much imagination to see many of the risks associated with providing family members, physicians, or insurance companies with such powerful technological capabilities. To some extent, heavy social, ethical, legal, and privacy concerns are hampering rapid progress in the field as technological development would allow.

For exactly these reasons, it is often very difficult or even impossible to observe primary users in action to help the analyst gain a sufficiently detailed understanding of the users' context, constraints, and work habits to generate useful user interface designs (Bennett et al., 2006). This is especially true in closed environments such as in hospitals and military contexts. For the same reasons as mentioned, researchers in medical environments are unlikely to be granted access to actual patient records or to live patient data even if this would not involve live observations. This means that many of the usual investigative and evaluative methods applied in user-centered design simply cannot be used. Requirements must be gathered from indirect data and secondary sources, and evaluation of successive prototypes must rely on artificial data and on quasi-laboratory studies before a clinical trial can be performed.

This chapter presents a case study reporting the challenges involved in the analysis, design, and evaluation of a system that tracks, records, and alerts medical staff about the state of vulnerable, seriously ill patients. The target population is nurses, physicians, and respiratory therapists caring for patients who are being discharged from an intensive care unit or other critical-care unit into a regular hospital ward. The system is about to undergo clinical trial; thus, it has not been implemented yet. Indeed, a *fully integrated system* allowing remote *continuous* monitoring of *multiple* patients' vital signs does not yet exist; the way in which this exacerbates the challenges of defining and capturing the users' needs as well as designing and evaluating the system is reported.

The next section describes the integrated system and the context in which it will be implemented, followed by a brief outline of the state of current physiological monitoring equipment. A brief summary of existing systems providing similar, albeit not integrated, capabilities is then presented. This is followed by a discussion of the challenges involved in conducting the user needs analysis in a closed environment. A method originally developed to generate future system users' requirements for novel mobile features to support feature bundling is outlined and amended to suit the present purposes. The challenges involved in the design of the various components of the present system are then outlined, followed by presentation of the specific requirements associated with the evaluation of the different system components and, finally, by a general conclusion.

THE INTEGRATED SYSTEM AND THE USAGE CONTEXT

The system described here began as a military-sponsored program within which the aim was to create a specification to form the basis for the development of a high-level proof-of-concept system capable of monitoring, recording, and storing several types of physiological data and then transmitting these in real-time to a central monitor. Upon successful completion of that project, the team was invited to further develop the system for use in a medical environment. **BRY**TECH, the local Ottawa company that is developing the present system, specializes in portable healthcare monitors and already has a working relationship with the Ottawa General Hospital where one of the company's patient vital-sign monitoring systems is already in use. The integrated system discussed here was conceived of when nurses and physicians using the existing system began to suggest desired modifications and additions that would increase its usefulness to them and their patients.

The integrated system comprises several components. A number of laptop base stations (central monitors) will be located around the hospital for convenient access by the key target user populations, including critical-care nurses, critical-care physicians, respiratory therapists, patients, and eventually—when the system is extended to enable monitoring of vulnerable patients in their homes—also their caregivers. A patient monitoring unit (PMU) connects sensors placed on the patient's skin with the system; it is either attached to the bed or, if the patient is ambulant, to a belt or a harness that can be worn comfortably under normal clothing. Members of the Rapid Assessment of Clinical Events (RACE) team—the specialist ICU nurses, critical-care physicians, and respiratory therapists—may be located anywhere in the hospital when an emergency occurs. Consequently, these professionals will be equipped with PDAs.

In the first instance, the system will be accessible via the central monitors and PDAs from anywhere in the hospital. Specialized hospital facilities such as emergency rooms (ER) resuscitation areas, Intensive Care Units, and recovery rooms typically provide advanced built-in monitoring capabilities. Dedicated, specialized staff provides one-on-one care. By contrast, the nurse-patient ratio in regular hospital wards ranges from 1:4 to 1:8 depending on the ward, time of day, acuity of illness, and nursing workload. Typically, the occupancy rate in Canadian ICUs, postoperative recovery rooms, and emergency departments exceed 95%. A review of monthly ICU occupancy rates of the Ottawa General Hospital over the past year reveals that rates always exceeded 95% and exceeded 100% more than half of the time. As a result, patients remain in the ER and recovery rooms for extended periods of time. It also means that many patients who are still vulnerable are discharged to the general wards even when the critical-care staff would have preferred them to remain in those critical-care units in which one-on-one monitoring is provided.

At best, patients' vital signs can be monitored on the ward every 2-4 hours, and even this frequency can be difficult to arrange. Failure to monitor patients more frequently can result in poor outcomes (e.g., Bedell, Deitz, Leeman, & Delbanco, 1991). Bedell et al.'s research suggested that approximately one in four cardiac arrests could have been avoided if stricter attention had been paid to a patient's signs and symptoms. It is a constant challenge, therefore, for hospital ERs and ICUs to balance insufficient numbers of medical staff with the need to monitor an increasing number of high-risk patients on a continuous basis.

Remote monitoring of patients in regular wards is currently not possible because the necessary equipment is typically hard-wired into the walls and/or the ceiling in the critical-care units. The system described here addresses this problem by providing mobile monitoring capabilities for patients on regular wards that is almost as good as the non-invasive monitoring in critical-care units. It also enables simultaneous monitoring of multiple patients as well as increasing the ability of ward staff to monitor patients' vital signs continually as is done in these specialized units. Many hospitals across Canada have developed a follow-up service upon discharge from the ICU (e.g., Bellomo et al., 2004; Salamonson, Kariyawasam, van Heere, & O'connor, 2001; Buist et al., 2002; Daly, Sidney, & Fatovich, 1998; Hourihan, Bishop, Hillman, Daffurn, & Lee, 1995) but the service does not exist to follow discharge from other critical-care units. The ubiquity of the present system overcomes this problem as well, thereby also addressing the frequent care gaps that still prevent staff from monitoring as many patients as closely as their condition would warrant, and that the staff would desire. The present system enables simultaneous monitoring of several vital signs in multiple patients who may be located anywhere in the hospital. It facilitates the rapid move of patients from one location to another and as well, decreases the need for high-cost ICU care for some of these patients. This capability alone is intended to enhance the hospital's ability to provide cost-effective, flexible options for monitoring vulnerable patients while in hospital, and eventually also to extend the benefits to high risk patients being discharged to their home. The prevention of medical complications will ultimately lead to fewer unplanned or urgent health interventions such as emergency visits or hospital re-admissions after discharge. In turn, this should reduce wait times in ERs and ICUs. The unique alarm systems integrated into the software are also expected to decrease nursing workload by drawing attention to those patients who need urgent attention.

CURRENT STATUS OF PHYSIOLOGICAL MONITORING

Several physiological monitoring technologies have been deployed in the area broadly referred to as telemedicine. Telemedicine is the use of telecommunications technology in diagnostic and therapeutic medicine. Several studies have assessed the impact of telemedicine in various aspects of patient care, including home care (Currell, Urquhart, Wainwright, & Lewis, 2000; Dale, Connor, & Tolley, 2003), ICU care (Becker, 2000; Rosenfeld et al., 2000), and teletriage (Lowery, 2001; Bunn, Byrne, & Kendall, 2004). Although these and other studies generally conclude that patients fare better in telemedicine than in conventional programs, most of the studies focus on intermittent monitoring of patient vital signs rather than on the continuous monitoring and data transmission required in the present case.

Of the few studies that do focus on continuous in-hospital patient monitoring—one of which collected vital signs data in a sample of 165 children during the night (Johnson, 1998; Lowery, 2001)—their authors have commented only on the feasibility of using the technology employed. Another study that did report monitoring data continuously from patients hospitalized with heart failure, however, had data that were transmitted only intermittently (Bondmass, Bolger, Castro, & Avitall, 1999). No study was found in which data were collected and transmitted continuously, and that enabled simultaneous monitoring of multiple patients. In addition to meeting several other stringent performance and display criteria, the system described in this chapter fulfills all of these requirements.

Physiological monitoring systems typically comprise patient-monitoring equipment located at the bedside and a central monitoring station. The patient is connected to the bedside monitor via cables and to the central monitoring equipment (generally located in the nurses' station) through another set of cables located in the walls and/or the ceiling. Both sets of cables usually employ up-to-date networking technology for gathering, distributing, and displaying data. However, the cables are usually dedicated to one particular function and are tied to the physical location in which they are installed. Transmission across standard Ethernet LANs such as those used in hospitals, normally requires additional coding because manufacturers tend to rely on different, non-standard communication protocols. Each bed requires its own wiring, making both expansion of the number of monitoring stations and the movement of patients to other wards cumbersome if not impossible, and necessitates the need for additional equipment. During transportation, the patient is disconnected from one system and reconnected to another. This can be problematic because the two systems are likely to have been manufactured by different vendors, thus, the data output, records, and displays may be incompatible with each other. In addition, data collected from portable equipment often cannot be transferred directly to the patient's record upon arrival at the destination ward or hospital. As a consequence, at least some real-time data are inevitably lost. Worse, if a medical incident occurs while the patient is in transit, paramedics may not be able to obtain expert advice and respond in a timely fashion because the data cannot be uploaded to the central monitoring system. Hospital physicians, therefore, are unable to evaluate the patient's condition and provide the necessary medical advice. In addition to data recording problems, portable monitoring systems are usually quite bulky, have limited functionality, utilize a large number of cables and heavy batteries as well as suffering from limited battery-life, and monitor displays are often difficult to view in bright light.

COMPETING TECHNOLOGIES AND PRODUCTS

Although several patient monitoring systems are available (e.g., LifeSync®, Lifeshirt®, AdvancedBPM®, Micropaq®), none of these are capable of monitoring multiple patients in multiple sites and providing multiple data types continuously and in real-time. For example, the LifeSync® is designed for high acuity settings and does enable wireless transmission of a patient's electrocardiogram (ECG) and respiration signals to a monitor within the hospital. Although this wireless system can provide continuous monitoring of up to 47 patients at a time, its range of transmission of signal information is limited to approximately 30 feet. The LifeShirt® continuously measures over 30 vital signs including cardio-pulmonary, blood pressure, blood oxygen, EEG, among others. In this system, data are stored on a data card which is carried in the belt or pocket. Although this design gets around the limitations of distance, the patient is responsible for uploading the data via the Internet for physicians to review and is not able to alert the physician of an imminent problem in real time. The AdvancedBPM® will apparently transmit BP and other vital signs using a microchip patch; however, this system is not yet currently available in the market. The closest competitor to the BRY-TECH RPM™ device is the Micropaq® which can monitor as many as 60 patients for up to 96 hours. However, the Micropaq® has limitations which preclude its use for the current situation. Monitoring is restricted to local in-room transmission but their system cannot support cross-floor mobility because it has a very short range—although it can continuously monitor the patient when the signal is out of the telemetry range. Consequently, real time problems that occur outside of the local range are not detected immediately thereby precluding emergency response. The Micropaq® also requires its own dedicated 802.11 network and cannot use the existing hospital network. This may result from regulatory issues rather than technical ones. A user needs analysis, discussed later, revealed that these systems did not meet the existing needs of medical staff.

GENERATING USER—AND SYSTEM REQUIREMENTS

User-centred design (UCD) advocates recommend integration of UCD and software engineering practices at all stages of hardware and software development. However, Mayhew (1999), who presents a thorough and detailed usability engineering lifecycle model states that although "Usability Engineering tasks parallel or overlap with traditional development tasks and are tightly intertwined with them, the exact overlapping of the two types of tasks is not completely clear." Indeed, published attempts to explicitly integrate usability engineering/UCD methodologies with software engineering practice have only come from the requirements engineering community (Sutcliffe, 2002).

The task of, and responsibility for, gathering, documenting, and translating systems requirements, therefore, is still a fuzzy affair with partial 'ownership' claimed by three communities: software engineering, requirements engineering, and usability engineering. With its emphasis on 'clean' coding, software engineers focus on "designing the thing right," whereas requirements engineers concentrate on "designing the right thing" (Boehm, 1981, cited in Sutcliffe, 2002). One could argue that the UCD/usability engineering community aims at achieving both by "designing the right thing right." Although requirements engineering (RE) attempts to explicitly integrate a user perspective into its processes, with few notable exceptions (Diaper, McKearney, & Hurne, 2000; Constantine & Lockwood, 1999) there is a tendency within the UCD community to develop usability engineering methodologies that, at best, align with software engineering and/or RE processes rather than aiming at integrating UCD with these. Importantly, Sutcliffe (2002) sees the goal of RE as proceeding from informal, fuzzy individual statements of requirements to a formal specification that can be understood by all stakeholders. By contrast, the goal of UCD activities during the user needs analysis (UNA) may be seen as gathering those "informal, fuzzy statements" from users/user groups representative of the target population(s), to verbalize collective user needs, distinguish these clearly from individual user wants, and eventually converting those statements into user interface prototypes.

The overlap between these communities notwithstanding, there is a scarcity of appropriate support methodologies specifically designed to guide the iterative data collection, analysis, presentation, and updating of findings that could bring all these three communities closer together and to clarify tasks and responsibilities. Because of the focus on challenges related to the HCI aspects of the present system, the further discussion on requirements focuses exclusively on the user needs analysis (Lindgaard, 1994). Information about the technical requirements is presented elsewhere (Lindgaard et al., 2007). A method for generating user requirements but which does not attempt to integrate UCD and RE is discussed next, followed by a description of the ways UNA was modified to suit the present purposes.

STRATEGIC USER NEEDS ANALYSIS (SUNA)

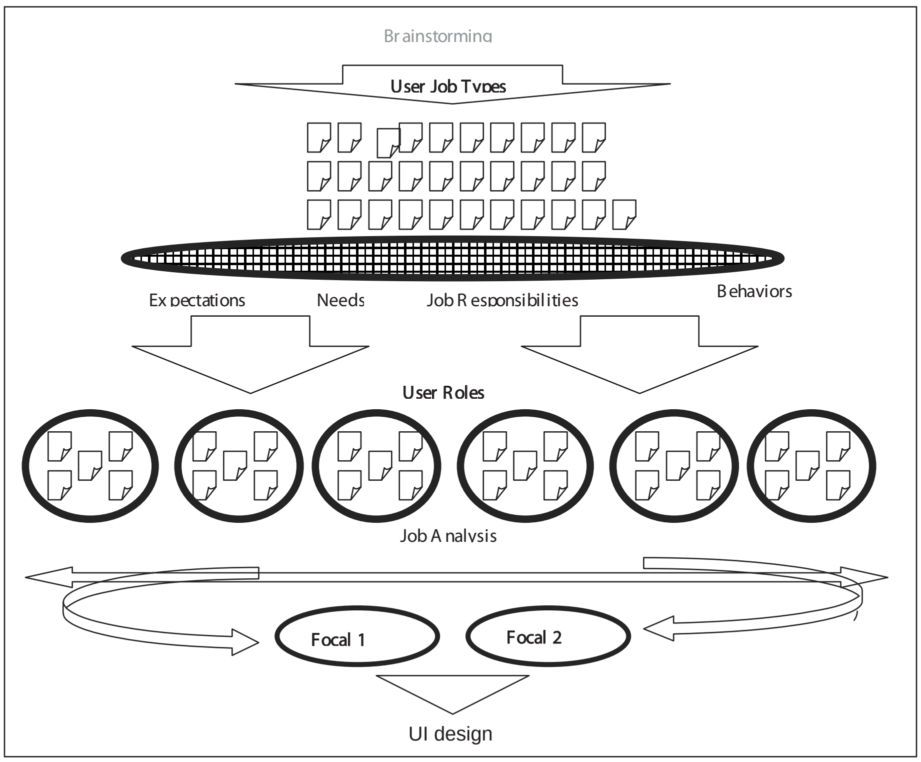
Strategic User Needs Analysis (SUNA) was specifically designed to predict the uptake of features in novel mobile devices (Narasimhan, 2001, 2004; Narasimhan & Lindgaard, 2002; Lindgaard & Narasimhan, 2007) in an effort to assist service providers to bundle features in ways that suited people working in different roles. Contrary to the approach commonly taken in UCD (e.g., Hackos & Redish, 1998; Beyer & Holtzblatt, 1998; Preece, Rogers & Sharp, 2002; Stone, Jarrett, Woodroffe, & Minocha, 2005) in which the UNA proceeds from a focus on individuals representing particular user groups, the starting point of SUNA is to develop an understanding of the roles and responsibilities of key user groups. Thus, rather than extrapolating from direct observations, interviews, job shadowing, and so forth, of individuals to the target user populations in a bottom-up fashion, SUNA takes one step back focusing instead on the higher-level user roles in a more top-down manner. This approach is consistent with Constantine and Lockwood's (1999) recommended 'usage-centred' (UC) design methodology. One advantage of using SUNA is that it avoids the potential problem of over-reliance on incomplete information obtained from individuals who may be less representative of the target population than analysts believe. In extreme cases, especially where only a small sample of future users is involved in the UNA, over-reliance on a bottom-up approach can lead to the development of applications that are quite inappropriate and which fail to support users and their tasks in the manner and to the extent intended. In addition, the focus on roles is likely to result in applications that are more readily and more widely generalizable than those generated from a bottom-up UNA. For these reasons, it was decided to revise SUNA to suit the present purpose and use the modified version to guide the user requirements capture as well as the early user interface (UI) design phases. The original SUNA and the modified approach are described next.

Steps in SUNA

The steps in the original SUNA method included a brainstorming session in which a wide range of job types was identified first. A group of users representing a variety of disciplines participated. After duplicates had been removed from the list of job types they were then grouped according to perceived expectations, needs, job responsibilities, and behaviors associated with each. As in Constantine and Lockwood's (1999) framework, 'expectations' referred to perceived device reliability and transmission speed, access to the network and user mobility. 'Needs' included issues such as receiving e-mail while on the move. 'Job responsibilities' were typical activities a person would be expected to perform in the particular user job type such as checking inventory or accessing a particular database. 'Behaviors' referred to perceived feature-usage patterns. The grouping of user job types continued until each was uniquely distinguishable from all the others. Job types were then combined into user roles of which six emerged from the analysis (see Figure 1).

User roles were subsequently verified by comparing the expectations, needs, relevant job responsibilities, and behaviors outlined in formal job classification documents obtained from the human resources departments of two large organizations. Two focal roles were then selected. Representatives of the two focal groups were involved in further interviews and journaling of their activities to verify the findings from the analysis and the combined findings formed the basis of the UI design.

Figure 1. Adapted from the original SUNA method (Narasimhan, 2001)



SUNA Modified

The high-level goals for this system were clear from the beginning. The hospital wanted a flexible, cost-effective, readily upgradeable solution that could enhance its ability to provide better monitoring of vulnerable patients in circumstances in which these patients were presently unmonitored due to technical or staffing problems. Stringent ethical and legislative patient and patient-record privacy standards and requirements had to be met. The nurses and physicians wanted better remote monitoring of more types of vital signs and of more patients—distributed across several hospital wards as well as in the patients' homes—without increasing their workload. They wanted to be notified reliably and immediately in case of a sudden change in a patient's condition, and for this notification to be clearly distinguishable from other existing alarms.

It was unnecessary to define user job types as it was already known who would be involved and what their jobs were. The user roles were also reasonably well defined *a priori* which made it unnecessary to verify the roles against job descriptions. It is, of course, possible that some of these may change as a consequence of introducing the system; consequently, the present data represent the project team's best guess. Before considering the expectations, needs, and so forth of the user roles, it was necessary to clearly identify the parts of the system with which the different primary and secondary users are likely to interact. These are presented in Table 1.

Primary users are those who interact directly with at least some parts of the system; secondary users typically receive reports or other outputs from the system. Secondary users may occasionally interact with some system components also but never with the full system (Lindgaard, 1994). Further inspection of Table 1 shows that the ICU nurses will be the main primary users together with the critical-care physicians and respiratory therapists who together form the RACE team. The ICU nurses take responsibility for the continual monitoring of patients recently transferred from the ICU to a regular hospital ward; members of the RACE team, however, will respond only to an alarm indicating a clinical emergency. Thus, the ICU nurses will be interacting with all the components of the system whereas the other RACE team members will predominantly use their PDA and the base station.

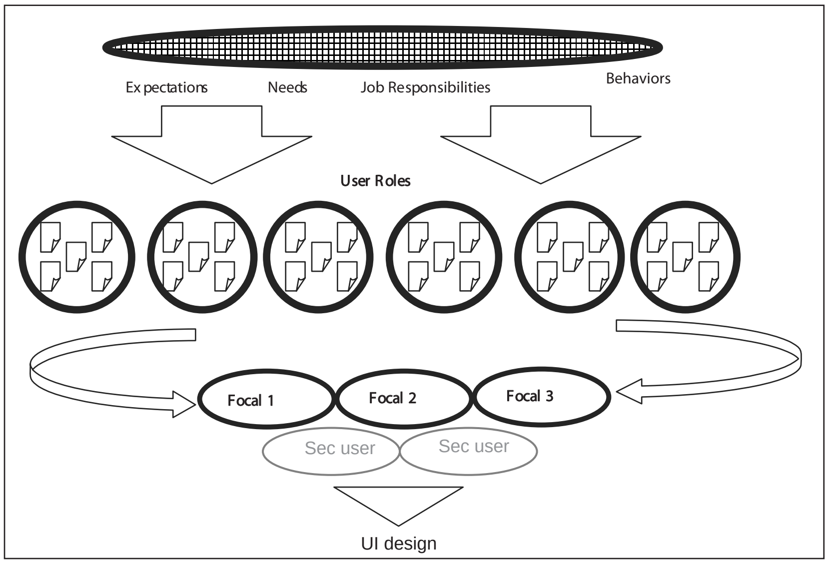
Patients and patient caregivers will be restricted to the patient monitoring unit and the harness and/or the belt; the same is true for ward nurses as well. Ambulant patients may be able to connect and disconnect themselves, for example, during bathing, whereas ward nurses need to be able to re-connect a patient to the system at any time. The extent to which the attending ward physicians will be inspecting records from the system is unknown at this stage.

The way and extent to which people in some of the roles may eventually interact with the system is still somewhat uncertain because these roles are yet to unfold with respect to the system. Subsequent analysis focused mainly on the needs of

Table 1. Interaction needs of primary users

User-type	User role	Central Monitor	PDA	PMU	Harness/belt
Primary	ICU nurses	✓	✓	✓	✓
	Critical-care physicians	✓	✓		
	Respiratory therapists		✓		
	Patients			✓	✓
	Patient caregivers			✓	✓
Secondary	Ward nurses			✓	✓
	Ward physicians	Possibly			

Figure 2. The modified version of SUNA used in the present system



members of the RACE team, namely ICU nurses, critical-care physicians, and respiratory therapists, although consideration was also given to the needs of patients and their caregivers.

The new monitoring system provides capabilities that are currently not available; consequently, it was necessary to first clarify the expectations, needs, job responsibilities, and behaviors associated with the primary (focal) user roles included in Table 1—particularly the ICU nurses and the RACE team members. The original SUNA described was modified to accommodate the secondary user (sec user) roles (see Figure 2). Three focal points are included to represent the ICU nurses, other RACE team members, and patients/caregivers. The outcome of modified SUNA analysis is shown in Table 2. It should be noted, however, that the issues identified here continued to be refined in subsequent participatory design meetings.

PARTICIPATORY DESIGN

At this point, the analysis and the UI design activities merged to become almost indistinguishable: Certain design decisions revealed gaps in the analysis and the refinements led to modified design decisions. The team concentrated on the UI for the central monitor. As indicated earlier, the primary user community was very small; the RACE team was already carrying out the most important role of acting quickly in emergencies, the users had prior experience with an earlier system designed by the same developers, and users were highly motivated to take part in shaping this evolving integrated system. Although most technology developments are no longer directed at supporting specific tasks for a well-defined community of work (Beck, 2002), the present UI design offered an opportunity for the team to refine the RPM™ system in a series of iterative participatory design (Schuler & Namioka, 1993) sessions. The developers translated decisions made in one session into high-level changes to the UI prototype in preparation for the next session. The final set of vital signs included in the central monitor UI comprise heart rate, ECG, respiration rate, temperature, non-invasive blood pressure (NIBP), and blood oxygen saturation rate. These are plotted graphically on the main screen on which the vital signs of multiple patients can be viewed simultaneously, and the data can be viewed comfortably from a distance. A variety of display configurations, which are changeable

Table 2. Results of modified SUNA analysis for primary and secondary users

User role	User type	Expectations	Needs	Job responsibilities	Behaviors
Primary	ICU nurse	Must be easy to use 100% reliability Easy switch between PDA and laptop Alarms audible immediately and anywhere	Monitor multiple patients simultaneously, continuously, and remotely Monitor multiple vital signs Audible alarms in any context Switch patient views quickly View both ECG and HR Move patients without disconnecting them from the system Review and edit patient information	Act quickly in clinical emergency Continually monitor vulnerable patients	Frequency of usage depends on state of recently discharged/ patients and number of clinical emergencies
	RACE team	Must be easy to use 100% reliability Alarms audible immediately and anywhere	Audible alarms in any context Switch patient views quickly View both ECG and HR	Act quickly in clinical emergency	Frequency of usage depends on number of clinical emergencies
	Patients/ Caregiver	Easy to connect/ disconnect Physically comfortable Easy to recognize when battery low	Lightweight belt/harness for patient unit Clear battery-state display Self-evident how to connect Easy to check connection		Usage only after discharge to home
Secondary	Ward nurse	Access to patient data	Read data Printout ECG File data	Monitor patient	Frequency of usage uncertain Likely to be infrequent
	Ward physician	Access to patient data	Read data online or printout	Ongoing responsibility for patient while hospitalized in ward	Frequency of usage uncertain Likely to be infrequent
Main Stakeholder	Hospital IT department	System must be easy to update and maintain 100% reliability No interference with existing networks Must meet existing standards			Frequency of usage uncertain Likely to be infrequent
	Hospital admin.	Must meet hospital policies on noise Must meet Health Canada and other standards			No interaction
	Hospital CFO	Must be cost-effective			No interaction
	Health Canada	Must comply with prescribed safety and other standards Must be clinically tested according to prescribed and approved protocol			No interaction

by a single click, enable the size of a particular patient display to be enlarged in order to keep that patient in focus.

The software tracks and records these five vital signs continually; the history of each vital sign can be viewed from another screen. Comments can be added to the patient file, but not deleted, at any time. The user can select a particular historical period to be inspected or they can view the entire patient history which is then compressed in the screen view. There are two alarm levels for which thresholds can be set for each individual patient and for each vital sign separately: 'amber alert' indicating that a particular value is above or below the threshold for the relevant patient albeit not dangerously so, and 'red alert' signaling a clinical emergency requiring immediate attention. Alarm states are shown visually on the main screen.

Even though members of the RACE team will be carrying PDAs to ensure they are notified immediately when a patient is in a state of clinical emergency (i.e., one or more values are in 'red alarm'), it is unclear if team members will rely on it for routine patient monitoring purposes. If they do, it is also not known how often, or under what conditions they will use it for data entry, and so forth. For that reason it was decided to equip the PDA with the entire functionality and with the same displays as the central monitor UI.

The main considerations in the RPM™ Patient Monitoring Unit design were ergonomic: it must be as small and as light as possible while also running for as long as possible on battery power, attaching the sensor electrodes to the PMU must be simple, erroneous or incomplete connections of sensors electrodes to the ports must be prevented at all costs, and the remaining battery power must be easily discernible.

The design of the harness was subject to similar considerations: it must be light and easy to fit correctly and adjust to different sizes, it must be comfortable to wear and remain in place even for prolonged periods of time, it must be easy to clean and of durable material, it must be unobtrusive when worn under normal clothing, free of restrictive wires and cords, and it must be socially acceptable.

EVALUATION

The evaluation of this system is probably the biggest challenge from a HCI point of view because, as was true for the UNA, it is not possible to obtain usability-related data from real patients in the context of the actual work, nor is it possible to observe the nurses in action. For example, it is not possible to obtain usability-related data from real patients in the context of the actual work, nor is it possible to observe the nurses in action. Lack of direct access compromises the ecological validity of several parts of the assessment for which alternative evaluation methods had to be devised.

The PMU and the Harness

The patient monitoring unit UI contains only slots for the sensor electrodes that will be attached to the patient. Although it was not possible to test the system on real patients, efforts were made to minimize its size, weight, and complexity based on knowledge of medical restrictions as well as physical and cognitive changes associated with aging. In addition, informal tests were conducted in-house of the PMU and harness, separately and together.

One member of the team was hooked up to the PDA, the central monitor, and the PMU to ensure that the components could talk to one another. The individual's BP, ECG, and SPO2 were monitored for more than 2 hours and data were recorded via the wireless network. One comment was that the noise of the automatic blood pressure check caught the participant off guard the first time, but he found that he was able to adapt to this quickly. The BP is a standard pressure cuff system but remotely controlled. The BRYtech system employs the Suntech Oscar2 unit adapted for remote control by the RPM system.

The waist belt and harness, in which ambulant patients will carry the PMU, were discussed and evaluated informally in several 2 hour sessions by two members of the project team, both adult males. Feedback led to a range of different designs. The resulting harness reportedly "felt fine" and was "comfortable," but occasionally it "feels like the equipment is sliding down." A belt hook may be needed to fix this.

At first, the shoulder harness (weighing 2-2 ½ lbs) seemed a little heavy. The battery pack was identified as the main contributor, weighing roughly 100 grams. The design currently uses off the shelf components which tend to be heavy. By changing to high density pcbs housing displacers which have all the functions in one chip, shorter cables in the PMU pouch for home use, and lighter albeit less durable plastic components for home use, the weight can be reduced considerably. The tradeoff is cost: This modification will cost $1-2 million which is infeasible for the time being. In the hospital, long cables and high grade metal parts instead of plastic are required but weight is less of a concern here because they are contained in the ICU pouch attached to the patient's bed.

The Central Monitor (Base Station)

The RPM™ central monitor was evaluated by a conventional heuristic evaluation and a user test with ICU nurses and physicians completing a set of typical user tasks. The heuristic evaluation uncovered some 46 issues. Contrary to the usual approach whereby issues detected in a heuristic evaluation are not fixed prior to the user test, most of the problems were fixed immediately because they were obviously problematic. Thus, 23 of the 46 problems were addressed immediately and prior to the user test, another eight items were placed on the 'to do' list and were addressed before testing the application, and the remaining 16 issues were resolved through discussion with the developer.

User tasks were generated from the heuristic evaluation and designed to ensure that all screens and typical tasks were represented in the task scenarios. One of the challenges in developing a credible cover story and task scenario is ensuring that all parties involved in the project (e.g., experts in medicine, usability, engineering and software development) have a shared understanding of what a cover story is, what a scenario is, and what these should or should not contain. After some discussion, the final cover story and task scenario were sufficiently credible to the 10 medical staff involved in the usability test. Some 75% of the tasks were completed without instruction or hint by all participants, and an additional 17% with just a single hint. Only 7% of tasks were not completed—all of which had been identified as issues still to be addressed. The average rating for *ease of use* was 70/100, and 77/100 for *user satisfaction*. For a system to be perfectly acceptable, one would expect both the ease of use and the satisfaction ratings to lie around 75/100. Thus, for the first phase alpha test, the central monitor base station application fared quite well.

The PDA

As was the case in the original SUNA method, it is not possible to obtain observational data directly when evaluating the PDA component in the present study. Indeed, it will probably not be possible to obtain any data before the system is implemented because the parameters of a clinical trial are so stringent that it will be impossible to burden the nurses with additional data requests. Instead, a small program that logs screen and feature activation and assigns a time stamp will be written. Output will record whenever the PDA is 'awakened,' what screens and which features are accessed, and the corresponding time stamp. At the end of each observation shift in the clinical trial, the logs will be used to jog the nurses' memory to describe where they were at the time and what prompted them to use the PDA in each episode logged. Observations will be conducted on each user for one shift at a time, sampling a total of 10 shifts from each nurse over a period of several months. This particular duration and frequency of observation are expected to suffice for the novelty effect to wear off and to learn more about how the PDA is actually used by the target population.

Alarms

In order to display the same amount of information on the PDA as on the central monitor screen, the Dell Axim X51-V PDA was chosen because it has a very high resolution. Unfortunately, this device does not have vibration capacity; current PDAs with vibration capability have poor visual resolution. The team opted for high resolution because of the need for visual clarity in viewing patients' records and setting/responding to alarm settings. Vibration-based alerts would be preferable in the hospital context because there are already numerous auditory alarms on existing equipment (Sanderson, 2006). The hospital requires that a slow pulse be used for less urgent and fast pulse for urgent alarms, using medium pitch. In order to ensure that the sound associated with this PDA meets the hospital requirements and that it is detectable, discriminable (Connell, Lui, Jarvis, & Watson, 2003), and recognizable, the range of existing sounds currently in use in the hospital will be tape recorded to identify possible choices for alarm tones. Upon selection of a sample of such alarm tones, the next challenge will then be to identify the best candidate. To do this, the ICU nurses will be requested to carry the PDAs during their normal working days. Different alarm tones presented at a range of intensities will be tested by randomly sending each tone to the PDAs and requiring the nurses to hit a particular button when they think they hear an 'alarm.' From the automatically generated data logs, the hits, misses, false positives, and false negatives will then be calculated and compared. As a control condition, a formal psychophysics experiment will also be carried out in the lab; the intensity of the same tones as tested in the field will be varied systematically with and without simultaneous background noise and with participants engaging in another independent task. The sound will be increased and decreased in small steps over several hundred trials. Nurses and physicians will only trust the system if an alarm can be heard without fail. Another important consideration that also affects the level of trust users will eventually have in the system concerns the reliability with which the alarm will go off when it is appropriate, while avoiding false alarms (Harel, 2006).

FUTURE USABILITY EVALUATIONS

Ideally, observations should be made of actual usage of all the interacting system components once the system is complete. Unfortunately, due to patient privacy and other ethical considerations, it will not be possible for the UI experts to obtain such data directly from situations involving patients. As is true for other situations in which novel technologies are introduced (e.g., other medical settings or in military outfits; Bennett et al., 2006), evaluation methods will need to be developed for obtaining reliable and accurate data from the medical personnel instead. These methods have not been developed as yet.

Finally, while the current PDA UI is virtually identical to the much larger central monitor UI, it will be very difficult if not impossible to ascertain the degree to which the same amount and types of data are really required for the PDA to maximize correct and speedy identification of critical information in patient emergencies. Some of the challenges here include the fact that users become accustomed to a particular UI even if it could be improved; user performance data may not be available to identify, for example, the average delay between a signal arriving at the PDA and the person responding to the emergency. Therefore, it will be extremely difficult to determine rules for switching an alarm off. For example, if the nurse is engaged in a task that must be finished before she can move to the patient in crisis and she has turned off the alarm, she may forget about the emergency by the time she is ready to move on to it. If the alarm is set to recur intermittently, this may be irritating if the nurse is on her way to the patient in crisis.

CONCLUSION

In conclusion, interface design and usability needs assessment in closed environments are both challenging and rewarding. Effective designs can make a large difference in the efficiency of the day-to-day operations of medical staff and impact enormously on patients' lives. Creative and alternative ways are needed to conduct ecologically valid evaluations of new interface designs without imposing excessive work demands on medical staff and also protecting patient record confidentiality. This case study identified a number of design challenges—most of which were overcome through discussion and active involvement of all members of the team (i.e., medical staff, hardware and software engineers, and usability experts). One solution used here was to employ and modify the Strategic User Needs Analysis approach to incorporate secondary as well as primary user roles. Next steps include automatic recording of logs on how the PDA is used, and how often, by members of the RACE team as a means of jogging memories for conducting post-use interviews during the clinical trial of the BRYtech RPM™ system.

REFERENCES

Anderson, K., & Dourish, P. (2005, July 22-27). Situated privacies: Do you know where your mother [trucker] is? In *Proceedings of the 11th International Conference on Human-Computer Interaction*, Las Vegas.

Beck, E. (2002). P for political—participation is not enough. *Scandinavian Journal of Information Systems*, *14*, 1-17.

Becker, C. (2000). Telemedicine system helps manage ICUs. *Modern Healthcare, 30*(37), 62.

Bedell, S. E., Deitz, D. C., Leeman, D., & Delbanco, T. L. (1991). Incidence and characteristics of preventable iatrogenic cardiac arrests. *Journal of American Medical Association*, *265*(21), 2815-2820.

Bellomo, R., Goldsmith, D., Uchino, S., Buckmaster, J., Hart, G., & Opdam, H. (2004). Prospective controlled trial of effect of medical emergency team on postoperative morbidity and mortality rates. *Critical Care Medicine, 32*(4), 916-21.

Bennett, G., Lindgaard, G., Tsuji, B., Connelly, K. H., Siek, K. A., & Khan, S. (2006). *Reality testing: HCI challenges in non-traditional environments.* Paper presented at Workshop CHI'06, Montreal.

Beyer, H., & Holtzblatt, K. (1998). *Contextual design: Defining customer-centred systems*. San Francisco: Morgan Kauffman.

Bondmass, M., Bolger, N., Castro, G., & Avitall, B. (1999). The effect of physiologic home monitoring and telemanagement on chronic heart failure outcomes. *The Internet Journal of Asthma, Allergy and Immunology, 3* (2).

Buist, M. D., Moore, G. E., Bernard, S. A., Waxman, B. P., Anderson, J. N., & Nguyen, T. V. (2002). Effects of a medical emergency team on reduction of incidence of and mortality from unexpected cardiac arrests in hospital: Preliminary study. *British Medical Journal***,** *324*(7334), 387-390.

Bunn, F., Byrne, G., & Kendall, S. (2004).Telephone consultation and triage: Effects on health care use and patient satisfaction. *Cochrane Database Systematic Reviews, 3.*

Connell, R., Lui, F., Jarvis, D., & Watson, M. (2003, July 14-18). The mapping of courses of action derived from cognitive work analysis to agent behaviours. In *Proceedings of the 2nd Autonomous Agents and Multiagent Systems Conference*, Melbourne, Australia.

Constantine, L., & Lockwood, L. (1999). *Software for use: A practical guide to the models and methods of usage-centered design*. New York: ACM.

Currell, R., Urquhart, C., Wainwright, P., & Lewis, R. (2000). Telemedicine versus face to face patient care: effects on professional practice and health care outcomes. *Cochrane Database of Systematic Reviews (2), CD 002098.*

Dale J, Connor S, & Tolley, K. (2003). An evaluation of the West Surrey telemedicine monitoring project. *Journal of Telemedicine & Telecare*, *9*(Suppl. 1), 39-41.

Daly, F. F., Sidney, K. L., & Fatovich, D. M. (1998). The medical emergency team (MET): A model for the district general hospital. *Australian and New Zealand Journal of Medicine*, *28*(6), 795-798.

Diaper, D., McKearney, S., & Hurne, J. (2000). Integrating task and data flow analyses using the pentanalysis technique. In J. Annett & N. A. Stanton (Eds.), *Task analysis* (pp. 25-54). London: Taylor & Francis.

Hackos, J. T., & Redish, J. C. (1988). *User and task analysis for interface design*. New York: John Wiley & Sons.

Harel, A. (2006). Alarm reliability: What if an alarm goes off and no one hears it? *User Experience*, *5*(3), 10-14.

Hourihan, F., Bishop, G., Hillman, K., Daffurn, K., & Lee, A. (1995). The medical emergency team: a new strategy to identify and intervene in high-risk patients. *Clinical Intensive Care*, 6, 269-272.

Johnson, P. (1998). The clinical and economic advantages of remote, community-based physiological assessment. *Journal of Telemedicine & Telecare*, *4*(Suppl. 1), 64-66.

Lindgaard, G. (1994). *Usability testing and system evaluation: A guide for designing useful computer systems.* London: Chapman & Hall.

Lindgaard, G., & Narasimhan, S. (in press). Factors influencing feature usage in work-related communication. *Journal of Behaviour & Information Technology.*

Lindgaard, G., Oakley, K., Kroeger, P., Miller, J., Bryenton, E., & Hébert, P. (2007). *Wireless technologies in a restricted environment: The challenging design of an innovative physiological monitoring system for high-risk patients.* Manuscript in preparation.

Lowery, C. R. (2001). *A retrospective study: Exploring selected characteristics of telenursing triage*. Unpublished doctoral dissertation, Wright-Patterson AFB, Ohio, Air Force Institute of Technology.

Mayhew, D.J. (1999). *The usability engineering lifecycle: A practitioner's handbook for user interface design*. San Francisco: Morgan Kaufmann.

Mynatt, E., Melenhorst, A., Fisk, A. D., & Rogers, W. (2004). Aware technologies for aging in place: Understanding user needs and attitudes. *Pervasive Computing*, 36-41.

Narasimhan, S. (2001). A user needs assessment for Personal Digital Assistants. Unpublished honours thesis, Carleton University, Ottawa, Canada.

Narasimhan, S. (2004). Activity theory and roles: Feature bundling for wireless technologies. Unpublished masters thesis, Carleton University, Ottawa, Canada.

Narasimhan, S., & Lindgaard, G. (2002). The importance of user roles in feature bundling decisions in wireless handheld devices. In J. Hammond, T. Gross, & J. Wesson (Eds.), *Usability gaining a competitive edge* (pp. 73-88). New York: Kluwer.

Preece, J., Rogers, Y., & Sharp, H. (2002). *Interaction design: Beyond human-computer interaction.* New York: John Wiley & Sons.

Rogers, Y. (2006). Moving on from Weiser's vision of calm computing: Engaging UbiComp experiences. In *Proceedings of Ubicomp 2006 (LNCS 4206)* (pp. 404 – 421). Berlin: Springer Verlag.

Rosenfeld, B. A., Dorman, T., Breslow, M. J., Pronovost, P., Jenckes, M., & Zhang, N. (2000). Intensive care unit telemedicine: alternate paradigm for providing continuous intensive care. *Critical Care Medicine*, *28*(12), 3925-31.

Salamonson,Y., Kariyawasam, A., van Heere, B., & O'Connor, C. (2001). The evolutionary process of medical emergency team (MET) implementation: Reduction in unanticipated ICU transfers. *Resuscitation*, *49*(2), 135-141.

Sanderson, P. (2006). Auditory displays in healthcare. *User Experience*, *5*(3), 19-22.

Schuler, D., & Namioka, A. (1993). *Participatory design: Principles and practices.* Hillsdale, NJ: Lawrence Erlbaum.

Stone, D., Jarrett, C., Woodroffe, M., & Minocha, S. (2005). *User interface design and evaluation*. New York: Morgan Kaufmann.

Sutcliffe, A. (2002). *User-centred requirements engineering: Theory and practice*. London: Springer Verlag.

Weiser, M. (1991). The computer for the 21st century. *Scientific American*, 94–104.

KEY TERMS

Heuristic Evaluation: An informal inspection method used in the field of human-computer interaction to identify any problems associated with the user interface (UI) design.

Heuristic Evaluation Violation: Occurs when one of the recognized usability principles (the "heuristics") is not complied with during typical user tasks. Such usability problems are categorized according to their estimated impact on user performance or acceptance and provide important feedback to the developers on the extent to which the interface is likely to be compatible with the intended users' needs and preferences.

Physiological Sensor Pod (SPOD): An external non-invasive sensing device that detects bodily signals such as blood pressure, heart rate, and respiration rate.

Usability: A term used to denote the ease with which people can employ a particular tool or other human-made object in order to achieve a particular goal effectively with efficiency and satisfaction. In the human-computer interaction context of this chapter, usability refers to the elegance and clarity with which the interaction with a Web site is designed.

Usability Test: The controlled experimental measurement of the ease with which a person can use a product or piece of software to achieve a specified goal in a particular context. Usability testing generally involves measuring how well test subjects respond in four areas: time, accuracy, recall, and emotional response.

User-Centered Design (UCD): Broadly defined as a design philosophy and a multi-stage problem solving process in which the needs, wants, and limitations of the end user of an interface is given extensive attention at each stage of the design process. This approach requires designers to analyze and foresee how users are likely to use an interface, as well as to test the validity of their assumptions in real world tests with actual users. This approach tries to optimize the user interface around how people can, want, or need to work, rather than forcing the users to change how they work to accommodate the system or function.

User Interface Design: The design of computers, gadgets, appliances, machines, mobile communication devices, software applications, and Web sites with the focus on the user's experience and interaction. Unlike traditional design where the goal is to make the object or application physically attractive, the goal of user interface design is to make the user's interaction experience as simple and intuitive as possible—what is often called user-centered design. In this chapter, the interface design is limited to mobile communication devices and associated software application.

User Needs Analysis: Provides a way to understand the gaps between what a person needs from a system and what the system actually provides.

Tether-Free Mobility Device: Refers to a wireless communication device that receives and transmits physiological data.

Chapter LXI
Designing a Ubiquitous Audio-Based Memory Aid

Shwetak N. Patel
Georgia Institute of Technology, USA

Khai N. Truong
University of Toronto, Canada

Gillian R. Hayes
Georgia Institute of Technology, USA

Giovanni Iachello
Georgia Institute of Technology, USA

Julie A. Kientz
Georgia Institute of Technology, USA

Gregory D. Abowd
Georgia Institute of Technology, USA

ABSTRACT

The ephemeral nature of sound can be problematic when people cannot recall something they heard. Motivated by everyday conversational breakdowns, a continuous, near-term audio buffering application named the Personal Audio Loop was designed to recover audio content from the recent past using the mobile phone platform. The investigation of the potential usefulness in everyday life, the level of ubiquity and usability required of the service, and the social and legal considerations for long-term adoption is presented. The methods used include a controlled laboratory study, in-situ diary and event-contingent experience sampling studies, examination of legislation, and deployment of the technology over several weeks.

INTRODUCTION

Everyday conversations fill our lives and we are all very familiar with the kinds of breakdowns suggested by these simple scenarios:

- You are in a conversation with a friend and one of you is interrupted. When the conversation resumes, neither of you remembers what you were talking about.
- You are at a social event and you are introduced to someone new. Minutes later, you have forgotten the person's name.

Current mobile phone technologies have the potential to preserve audio to enable people to recover needed information in many similar situations. Thus, the Personal Audio Loop (PAL) is the product of a multi-year design and research project focused on developing such capabilities. PAL is an automatic, audio-based memory aid that relies on audio capture technologies and the remarkable human ability to recover information quickly with only a few bursts of the audio from that memory.

PAL represents a significant case study for the design of new mobile and ubiquitous technologies. PAL is exceptional in its design and in its use because it required emerging technologies and unique usage models. Additionally, it is a personal technology that makes use of public information: the audio in the environment within earshot of an individual. In this chapter, the innovations required for this design case are explored. A mixed technological and human-centered approach was necessary to produce a near-term audio service that could survive research in the natural environment of people's everyday lives.

The design process addressed multiple questions of technical and human significance:

- **Usefulness:** Though motivated by observations from everyday life, how often and in what situations do people actually need a near-term audio memory aid?
- **Ubiquity:** What parameters of this service would make it available every place and every time it is needed?
- **Usability:** How should the service deliver functionality to maximize its benefit and minimize its distraction?
- **Social and legal considerations:** What aspects of society may influence the uses and cultural practices surrounding an audio recording application for everyday life?

PAL and technologies like it are becoming increasingly common, but at the time of its inception, the design was arguably outside the realm of a typical person's experience. Therefore, two essential requirements of the design process were to allow potential users to interact with a working prototype to have a sense of the capabilities and to answer engineering questions, including important architectural considerations.

From a technical perspective, there are several options for designing an audio-based memory aid to provide the required capabilities. All of the designs considered during this process reflected the same basic notion of replaying a buffer of recently recorded audio, but early prototypes varied in terms of the nature of recording and playback capabilities. A fully distributed system assumes an environment equipped with microphones, speakers, and interface controls to maximize opportunities for recording and playback wherever and whenever needed. A fully localized solution provides recording and playback in an all-in-one package carried wherever needed. A hybrid solution might delegate the recording in the environment and use a handheld device that receives streamed audio from a central repository for playback.

The complete design process for the Personal Audio Loop, a solution for an audio-based near term memory aid that addresses the technical concerns of an interesting capture and access application and answers questions from the four categories described, is presented. The process involved a series of formative studies that led to the design of a self-contained service integrated into a commercial mobile phone handset, summative studies that explored real life uses for primary users, and situated imagined uses that considered secondary stakeholders.

BACKGROUND AND RELATED WORK

Similar Audio-Based Memory Applications and Technologies

Near-term capture and access applications that provide audio reminder services have been previously explored in the office as well as for telephone conversations. Xcapture, originally built to provide a "digital tape loop" of a single office, could also provide short-term auditory memory of telephone conversations (5 to 15 minutes long) (Hindus & Schmandt, 1992). Although the system was designed for use in a setting where social protocol allows recording, the authors recognized the privacy issues of subsequent use of archived recordings and suggested that social expectations change with use. In MERL's real-time audio buffering technique, captured audio persists for the duration of that phone conversation (Deitz & Yerazunis, 2001). During the course of the conversation, a user may tap the phone against the ear to move backwards in the audio and to replay any portion of the discussion. This system does not store conversations and could arguably pass legislative tests and be socially acceptable. On the other hand, a reminder application can be created to preserve the audio for long-term use. In their investigation of an audio-based personal memory aid, Vemuri, Schmandt, Bender, Tellex, & Lassey (2004) created a mobile application for the iPaq PDA that constantly records, analyzes, and indexes audio to identify the best memory triggers for events from daily life. The application stores short audio clips, which it has determined to be good memory triggers for an event based on an analysis of the audio for key features such as the speaker and the tone of the discussion. Additionally, the system performs speech recognition to allow the user to perform keyword searches over the captured content at a much later time.

Relevant Legal Cases and Policies

Legal cases over the past two decades have exposed the contrasting requirements and balances of privacy and utility for recording applications. The experience in the fields of surveillance in public spaces and of the privacy of private communications is drawn from. Among other sources, European Directive 95/46/EC opinions and rulings by various EU Data Protection Authorities (DPAs) (British Institute of Comparative Law, 2003; European Commission, 2004), and several U.S. Supreme Court[1] rulings were considered.

Despite the ongoing debate stressing the differences between the United States and Europe regarding privacy, legislation regulating the recording of communications by electronic means is remarkably similar. The main items are the U.S. Electronic Communications Privacy Act (ECPA) of 1986 and European Directives 2002/58/EC and 95/46/EC. ECPA regulates wiretap and surveillance and applies to any electronic recording device and conversations ("oral communication") between two persons "exhibiting an expectation that such communication is not subject to interception," even if the conversations were not transmitted through a telecommunications network. European Directive 2002/58/EC covers only personal conversations transmitted over public telecommunication networks. However, Directive 95/46/EC applies to any personally identifiable information which includes recorded voice conversations, according to multiple opinions by European national data protection authorities. Directive 95/46/EC requires a proportionality assessment between potential harm and benefits; however, the personal character of the application might exempt users from many provisions, including informed consent.

FORMATIVE EVALUATIONS OF PAL

Based on early interviews with potential users and intuition, it was determined that the platform for PAL would need to be mobile, computationally powerful, include a robust development environment, and feature easily accessible interaction techniques and an external or attachable microphone. The mobility, ubiquity, and performance of mobile phones make them an appealing platform for this application but only certain phones support the required capabilities. The chosen Motorola iDEN i730 (Figure 1) clamshell phone features the J2ME programming environment, an external microphone, and buttons accessible while the device is closed and worn while clipped to a belt on in a pocket. The microphone is capable of recording voices in a small room with the phone open or closed in a shirt pocket or attached to a belt, with higher quality than most PDAs. Questions about the feasibility of the mobile phone as a platform as well as about the nature of situations that would require use of PAL remained. Thus, two formative studies were designed to answer these questions: a laboratory study designed to answer the questions of usability of the interface and a diary study to characterize the frequency and situations of use in everyday life, the full details of which are available elsewhere (Hayes et al., 2004).

Laboratory Study: Developing a Usable Mobile Interface

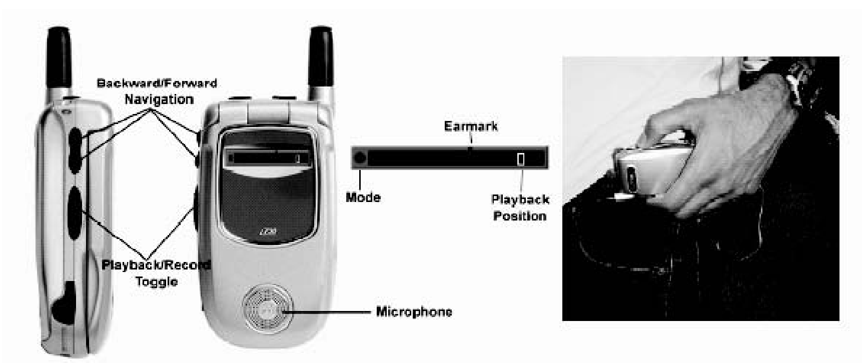
Mobile phone applications can be difficult to use. The limited display size and paucity of buttons or other interaction methods mean that creative designs are necessary to create feature-rich applications. PAL required several features including the need for quick, unobtrusive access and interaction techniques that enable a human to find relevant moments in a stream of audio. If PAL was too difficult or awkward to use or took too long to locate important audio, it would likely not be less useful for a person.

The Initial Prototype

The laboratory study necessitated a working prototype capable of testing the phone as an interface to the recent audio recorded in a person's surrounding. In its normal operating mode, the implementation of PAL continuously records audio from the user's environment. Audio older than the buffer length (in the initial prototype, 15 minutes) is deleted automatically. Recording automatically halts when the user answers or makes a call. Five buttons are available on the outside to accommodate interactions while the phone is closed (Figure 1). PAL provides simple audio navigation features (e.g., rewind), informed both by previous research on skimming (Arons, 1993) as well as by commercial video recording services like Tivo™. PAL includes a simple timeline visualization on the exterior LCD of the handset indicating application status (recording, playback, and direction of navigation) as well as the playback position in the audio buffer relative to the current time (the right edge of the timeline).

Technological limitations and engineering problems remaining to be solved meant that a fully functional prototype would take several months. In addition, developing applications for

Figure 1. Left: The Motorola i730 handset used for PAL. Three buttons control navigation and record/ playback mode. A timeline on the face of the phone indicates mode and relative place in the buffer. Right: Example placement of phone when using PAL.



a mobile phone is a lengthier process compared to traditional desktops. However, to test the usability of the design before exerting such a programming effort was wanted. As a result, a "tethered" version of PAL specifically for this controlled study (see Figure 2) was built. For the tethered version, a Motorola i730 mobile handset was connected to a laptop computer through a serial connection. The laptop handled the audio recording and playback using the Java Media Framework but the visualization and interaction with the application was entirely through the i730 handset. The button events were sent to the laptop through the serial link and the visualization was updated on the handset as events were sent back from the PAL audio engine running on the laptop. This architecture provided a quick and easy prototyping platform for the design. It also provided a way to easily log the user events as the user interacted with the PAL application. Without having to worry about the technical limitation of the mobile handset, a prototype which was indistinguishable from the actual wearable version was able to be quickly created. This early prototype allowed the gathering of quantitative performance metrics as well as user reactions to the device.

Method

The laboratory study included 18 participants: students and faculty from the Georgia Institute of Technology. Participants ranged in age at the time of participation from 18 to 50 and included five females. Participants had a background in Human-Computer Interaction and usability, and were chosen explicitly with the intent of examining heuristics such as the mapping of buttons to functionality and the quality of the visualization. Participants' experience with mobile phones ranged from 7 years of consistent use to no experience at all (7 participants). The prototype was demonstrated, encouraging participants to examine the device and ask questions until they expressed comfort with its functions.

Figure 2. Initial prototype of PAL tethered to a laptop used in a controlled laboratory study

PAL's intended use involves the replay of audio for which the user was present initially. The controlled study, designed to mimic this scenario, included a scripted dialog of 5 minutes. In this script, the participants asked researchers predetermined questions and researchers replied with the same answer for every participant. The script purposely involved a large amount of detail to increase the likelihood that participants could not recall the answers to all questions from memory. After completing the dialog, the researchers who had been participating in the dialog removed the script and asked the participants a series of questions about the information they just heard. Although the researchers noted whether the participants remembered the information without use of PAL, they asked every participant to find and play every answer. Participants were encouraged to "think aloud" as they used the prototype and the researchers timed how long it took an individual to find the answer, theorizing that this first time use while discussing their actions would be a worst case timing for most users. Participants answered seven questions, the first two being practice questions not used for computing timing results. An exit survey and semi-structured interview provided further evaluation of the interface by users and of their envisioned need for this kind of service.

Results

After a short demonstration, all participants were able to navigate the audio well enough to answer the questions. The exit survey had users answer questions and rate various features on a Likert scale from 1 to 7, ranging from 1 being "Strongly Disagree" and 7 being "Strongly Agree". According to the exit survey, participants agreed that the device was easy to use with one hand and small enough to carry at all times. They could clearly understand the audio even in its highly compressed form.

With an audio buffer of 15 minutes, participants required an average of 34.8 seconds to find responses for questions that were known to be in the in the recorded audio while talking aloud about their actions. Participants reported the visualization was somewhat helpful in accomplishing the task but not overwhelmingly so. Thirteen of eighteen participants actually used PAL without the visualization, preferring an eyes-free interaction.

Although inquiring about privacy was not a goal of the controlled usability study, 10 participants raised spontaneous concerns regarding the social acceptability of a continuously recording system. The most common sentiment expressed indicated that participants were less concerned about recording their own voice than those of their conversation partners.

Diary Study: Determining the Usefulness of PAL

The laboratory study showed the feasibility and usability of PAL on a mobile phone but it did not inform about the overall usefulness in everyday life. A diary study was undertaken to explore the extent to which a near-term audio reminder service was needed, looking for frequency and characteristics of potential use. Diary studies balance the ecological validity of gathering such data in situ against interruption of everyday activity flow caused by recording personal observations, particularly in mobile settings (Czerwinski, Horvitz, & Wilhite, 2004). Specific information relating to social context including privacy concerns in the diary entries and during the follow-up interviews was asked for.

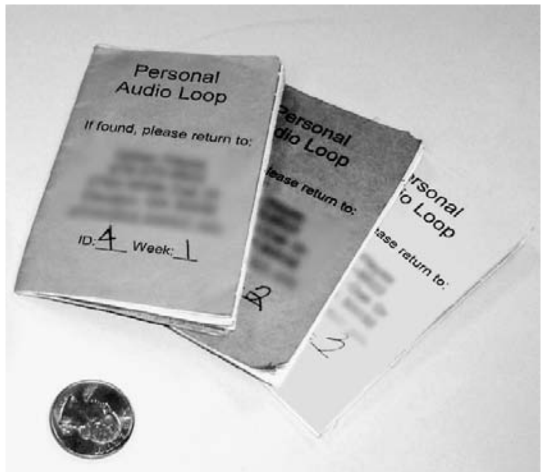
At the time of the diary study, a working version of PAL on a mobile phone was available. However, it was not ready for full deployment because of performance and reliability problems. The prototype was more than adequate for the purposes of demonstrating the system to participants. The diary study allowed the development of PAL to continue while still providing important development feedback.

Method

Twelve experienced mobile phone users (5 female, 7 male, ranging in age from 22 to 60 years) participated in the study. Participants' occupations spanned a spectrum of domains, including a psychologist, a finance manager, a realtor, a car dealer, a consultant, a professor, and a full-time homemaker. A working version of PAL was demonstrated to the participants. Then they were asked to carry small pocket-sized diaries and record an entry for each incident during the following week when they would have needed or liked to use the PAL service. Each page of the diary contained a simple form to complete for the potential instance of use, streamlined after an initial trial period. Each form in the diary included space for describing the content of the audio to retrieve, when and where the incident occurred, and whether any persons unrelated to the conversation were nearby. Participants also estimated how far in the past the salient audio content was and rated how important it was to retrieve that information. Figure 3 shows an example of an incident survey.

At the end of each week the diaries were collected from the participants and semi-structured interviews were conducted to examine a subset of entries in detail, up to six diary entries per participant per week. The detailed questions probed issues such as the kinds of information being sought, the distance of unrelated third parties from the participants, and the participants' assessments of the social appropriateness of using the device in the specific contexts. Each participant who chose to continue for another week was each then given a new empty diary to again record incidents. Participants could continue with the diary for up to 3 weeks. At the end of the study, semi-structured interviews with all participants were conducted.

Figure 3. Left: Sample diary entry; Right: Examples of pocket size diaries used in the study



The weekly and summary interviews allowed the clarification of misunderstandings in the entries as well as probing particular issues that were more difficult to garner from the diary form factor.

Results

Twelve people participated in the first week, eleven continued for the second, and eight in the third, for a total of 31 participant weeks and 109 incident reports. On average, participants reported more than three incidents per week, most of which referred to data that was less than an hour old with only 6% from over a day prior. Of the incidents reported, most occurred in public or semi-public places (defined as schools, workplaces, etc.) with less than a third in private space (predominantly car and home). In almost half of the incidents, people unrelated to the audio they wished to retrieve (e.g., other customers in a restaurant) were present. During follow up interviews, participants asserted that they would not have felt rude towards their communication partner using PAL in most of the situations. During the second and third weeks, participants were questioned about their reactions if their partners objected. Participants stated that such an objection would be "not likely" in almost every case and indicated that they would not have complied with the objection, had there been one, in three fourths of the questions. When asked how far away they would like PAL to record, two-thirds chose within a small room (10 feet) and only one individual requested a large radius, reporting that he is "just nosy." Generally, participants were willing to take the few minutes that would likely be required to access the data with willingness to spend time finding information roughly correlating to the importance of the information.

SUMMATIVE EVALUATIONS OF PAL

Deployment Study: In Situ Evaluation of Real Life PAL Use

The formative studies of PAL indicated how people might use PAL but it was difficult to know if these predictions were true without testing the service in the "real world." Thus, a robust version of PAL was developed on a mobile phone, which was informed by the results of the laboratory and diary studies. Having a fully functional system on the common form factor of a mobile phone allowed the exploring of how people would use such a system were it to be widely available.

The Final Prototype

The PAL application runs continuously on the i730 handset in one of three modes. The default mode is recording but that mode can be switched to either pause or playback. Due to the phone's performance limitations, there is some latency in the overall interaction including an 800-1000 ms delay for playback to actually begin after seeking. The laboratory study indicated this delay was still within a usable range. The phone that was used has enough memory to support up to 3 hours of audio; considering silence periods, the phone could provide recall functions for considerably longer. For the initial deployment, an arbitrary limit of 15 minutes for the recorded audio buffer was chosen although the application allows users to set their own buffer length. The i730 features a high quality built-in microphone and an external speaker. The external microphone is sensitive enough to pick up voices in a meeting room with the phone closed and in a shirt pocket. This range was sufficient based on the situations found in the diary study. The PAL application runs over 12 hours on a fully charged 1400 mAh lithium-ion battery with it continually recording and an average of 15 seconds of playback every hour.

The PAL application runs with the phone shell "flip" closed and the user interacts with the three buttons on the left side of the handset. PAL can be operated with the handset closed for complete heads-up operation, especially when worn on belt clip with a hands free ear bud and microphone; in this configuration, the external display provides status information. When the user opens the phone flip either to answer a call or dial a call, the recording pauses. In this mode it is also possible to browse the recorded audio and the internal display provides information to facilitate skimming. The users can skim backward, forward, and toggle between record, pause, and playback modes. In playback mode, quick and simple navigation through the audio stream is important. PAL is controlled using the available side buttons of the handset and a graphical display on the external panel for quick visual feedback. When a user needs to re-listen to part of a conversation, the small buttons are used to skip backward or forward a fixed amount of time in the recorded audio stream. This simple forward and backward jumping allows a user to skim a conversation to find a portion that reminds them of some salient feature of a previous conversation. The prototype defaults to a skip backward of 30 seconds and a skip forward of 10 seconds based on the controlled study. Significantly different forward and backward jump times were purposely chosen because similar forward and backward times would cause too little net change in skim mode when alternately using forward and backward.

It was found that different people have a different conception of which of the two navigation buttons is forward and which is back. As the orientation of the phone changes, these mappings also change. To alleviate this problem the notion of smart buttons, where the user is never wrong, was introduced. Each time the user initiates skim or playback of the recording, the only logical direction is backwards. The application makes whichever button the user presses first, the back button. The opposite button becomes forward. After the user goes back to record mode, the mapping resets. The next time the user initiates playback mode, again, the first button selected becomes the back button.

Operating on the metaphor of a running timeline, the display shows the full 15 minute audio buffer as a horizontal bar. There are two moving parts. The first is the buffer (dark region of the horizontal bar) that moves during recording to show that audio that has been recorded. After 15 minutes or the specific buffer length, the entire bar is filled in, which will be the case most of the time. The second movable part is a bead that shows the playback position relative to the recording position. By looking at the display, the user can position the playback point relative to the current recording point. In addition, users can use the visualization to roughly locate recordings.

To aid human's inaccurate recollections (e.g., when something occurred 5 minutes ago compared to 10 minutes ago), additional context information on the visualization is provided. One feature shows the silent time resulting from pauses during recording on the visualization. The recording stops when the user goes into playback mode, when a call is placed or received, or when the user explicitly pauses the recording. Since the timeline is a temporal representation, a dead space placeholder is provided (indicated by white space) to indicate where the mentioned situations occur relative to the current audio buffer. The white space is skipped when skimming. However, there is an important tradeoff to note with this scheme. Consider someone who talks for over 15 minutes on the phone or re-listens for a significant amount of time; most of the buffer would just be filled with white space. Therefore, an option for PAL to record either the last 15 absolute minutes or the last 15 actual audio minutes is provided. Finally, a marking feature where users tag important points in the audio loop is provided. This could be the result of something heard during record or playback mode. For example, if the user receives directions, they may place a bookmark at the time of recording because they know they will need to go back and access it later. To bookmark, the user simply taps the large button to indicate a tag point and a vertical tick mark appears on the visualization. The actual mark is placed 5 seconds before a button press event. In playback mode, holding the navigation button jumps between markers.

Method

A working version of the application was deployed to four of the diary study participants for 7 weeks. Furthermore, four members of the research team also used PAL for over 2 months, one as long as 4 months. During the first four days of the deployment, participants were asked to carry a diary to note their uses of the device. Throughout the deployment of the technology, their uses were also logged. By deploying the devices to even a small number of users, both expected and emergent uses could be observed and greater understanding about the dependency users might develop on the service could be gained.

Results

On average, participants used the device more than twice a week. Although this average is lower than what was indicated by the diary study, participants also reported an average of 1.5 incidents that they thought about using the device and chose not to use it. In one case, the user's conversation partner recovered the information before the user was able to try with PAL. In all other cases, the reason not to use PAL was reported as forgetting it was available. Interviews with the users since this initial probe indicate that ordinary use subsequently remained fairly consistent with the rate observed in the first 4 days and that the frequency of use for exploring the application or showing it to others has decreased substantially. Overall, satisfaction as reported through qualitative interviews has been high. All four users requested to continue using the devices after the first 4 days and reported that they believed they would use them more, over time. Each user changed the buffer length (ranging from 10 minutes to 60), the initial jump backward (ranging from 15 seconds to 60), or both. Users expressed that configuring the application was important and one user even indicated that he changes the buffer length depending on the situation he is about to encounter.

In the initial 4 day probe, the most frequently reported situation for use was to remember forgotten details (60%). Other unexpected situations have also been reported in the following weeks. Specifically, users have been employing PAL as an instructional aid. For example, one user recorded conversations with customers and then replayed them for an employee in training. Another recorded a negotiation with an employee to replay it for another, again as part of job training. One user regularly used the PAL service to augment her medical journal. Her physician had asked her to record data about symptoms. Her job required a great deal of driving, however, which made it difficult for her to record these symptoms by hand. Instead, she chose to speak them aloud when she could not write them down at that moment. She would then pause the recording long enough to replay it, when she could write in her health journal.

All of the users expected the service to be available, reporting that they would choose not to write information down when it is already spoken aloud. Ultimately, this expectation of availability became a significant hurdle to long-term adoption. No user continued to use PAL for more than 6 months. The primary reasons for discontinuing use were: (1) difficulties encountered when the operation of the device as a phone and the device as PAL interfered with one another and (2) annoyance with carrying two phones in the case of those users who were still carrying their personal mobile phones as well as the PAL-enabled phone.

Social contract issues recurred more often in the case of the real deployment than the results of the diary study had revealed. Users expressed that conversation partners aware of the device sometimes responded negatively initially, but relaxed after the application and its buffering and discarding functions were explained. Interestingly, all four users reported informing new conversation partners about PAL less frequently as time went by. After several weeks, users had almost stopped alerting conversation partners altogether. They would use their own judgment of the situation to decide if it was acceptable to record. They did, in fact, sometimes choose not to record despite reporting wanting to at those times. In both situations in which they ultimately chose to record and those in which they turned PAL off, participants reported deciding not to alert people to the possibility of the recording. As frequently as users reported negative social repercussions from PAL, they also reported positive cooperative uses of the device. For example, one user's wife consistently used PAL on his device by walking near him and speaking when she needed to remember something. Despite the in-depth usage information that was able to be obtained from this deployment of the service, the choice on the part of the users to avoid discussing PAL when they thought it might be confrontational left no in-depth understanding of the experience of the conversation partners and others nearby.

Contextualized Survey Study: The Experience of the Conversation Partners

Motivated by the results of the diary, laboratory, and deployment studies as well as available legislation and instincts about societal norms, it was chosen that the problem of PAL be explored further, from the standpoint of the conversation partners. As uncovered in the first three studies, primary stakeholders sometimes may have legitimate interests in using PAL, for example, due to a memory dysfunction or simply because of cognitive load. This interest may be opposed to that of secondary stakeholders or third parties (who might not want to be recorded, even if only temporarily, or who might want further explanation of the service when the user may find taking time for this explanation to be inconvenient or even impossible in the particular situation). Thus, there is a significant question of whose interests should prevail. A second question considered use of PAL relating to the proportion of individuals opposed to the application. If only a small minority of secondary stakeholders and third parties oppose PAL and the vast majority does not care, should the contrary minority be yielded to and a large market potential be curtailed?

These issues echo the classical ethical debate between utilitarianism and normativism in the context of privacy and technology (Terrell & Jacobs, 2002). It was chosen to explore these issues using the concept of proportionality (Iachello, 2006) borrowed from the legal community. To make determinations about the risk and benefit tradeoffs, the following needed to be understood:

1. To what degree, and in which situations, secondary stakeholders are most likely to object to the use of a device that can potentially cause the recording of their conversation (i.e., are objections unqualified or do they depend on the location, the topic, the identity of the conversation partner, or on the perceived confidentiality of the conversation?); and
2. What application parameters (e.g., retention time) can be adjusted to meet a compromise

between the interest of the primary users and conversation partners.

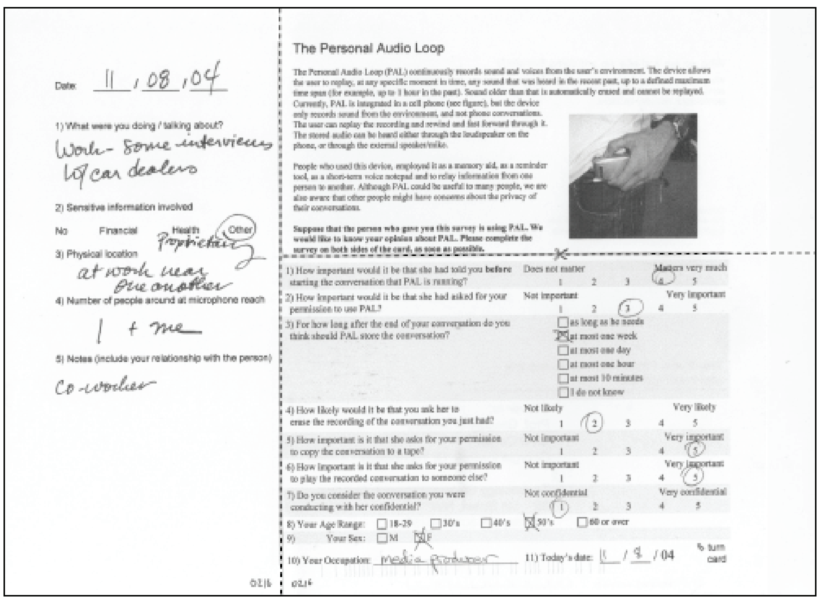
As stated in the section on the deployment of the technology, users might have been frustrated with the limitations of the prototype and thus, did not participate in a long-term study of the service. Rather, an event-contingent experience sampling procedure targeted at the conversation partners of PAL users was designed (Iachello, Truong, Abowd, Hayes, & Stevens, 2006). Experience sampling is a technique that has been in use since the 1970's to measure user responses in everyday settings to various stimuli (Scollon, Kim-Prieto, & Diener, 2003). Wheeler and Reis define event-contingent sampling procedures as those initiated by the occurrence of a specific event, in this case a conversation (Wheeler & Reis, 1991).

Method

A group of individuals whom are called "proxies" acted as substitutes or probes for the researchers in conducting the study. As the proxies went about their daily activities, they asked people older than 18, acquaintances as well as unknowns, at the end of their normal conversations, to read a description of PAL and to fill out an anonymous survey about the application. The proxies were instructed to hand out the surveys only if the conversation exceeded 3–4 sentences—weeding out very short interactions such as "thank you" and "excuse me." Also, the survey was only submitted once to the same individual to avoid respondent bias. The survey was administered in various locations in the United States.

The goal of the survey was to link the questioning to real life experiences as closely as possible. Thus, participants were asked to imagine that the person with whom they had just been speaking was using PAL. The survey included two parts, linked by a unique number (Figure 4). The investigator completed the first part (on the left) with information about place, participants, and the activity being achieved with the conversation, as salient elements of the social setting (Goffman, 1966). The investigator then detached the portion on the right and gave it to the participant along

Figure 4. Example of "proxy" survey. The survey form is divided in three parts, here shown after being reassembled. The left side is filled out by the researcher. The right side is given to the participant. It contains a description of the application and the survey.



Date: 11 / 08 / 04

1) What were you doing / talking about?

2) Sensitive information involved

No Financial Health Other

3) Physical location

4) Number of people around at microphone reach

5) Notes (include your relationship with the person)

The Personal Audio Loop

The Personal Audio Loop (PAL) continuously records sound and voices from the user's environment. The device allows the user to replay, at any specific moment in time, any sound that was heard in the recent past, up to a defined maximum time span (for example, up to 1 hour in the past). Sound older than that is automatically erased and cannot be replayed. Currently, PAL is integrated in a cell phone (see figure), but the device only records sound from the environment, and not phone conversations. The user can replay the recording and rewind and fast forward through it. The stored audio can be heard either through the loudspeaker on the phone, or through the external speaker/mike.

People who used this device, employed it as a memory aid, as a reminder tool, as a short-term voice notepad and to relay information from one person to another. Although PAL could be useful to many people, we are also aware that other people might have concerns about the privacy of their conversations.

Suppose that the person who gave you this survey is using PAL. We would like to know your opinion about PAL. Please complete the survey on both sides of the card, as soon as possible.

1) How important would it be that she had told you **before** starting the conversation that PAL is running? Does not matter 1 2 3 4 5 Matters very much

2) How important would it be that she had asked for your permission to use PAL? Not important 1 2 3 4 5 Very important

3) For how long after the end of your conversation do you think should PAL store the conversation?
- as long as he needs
- at most one week
- at most one day
- at most one hour
- at most 10 minutes
- I do not know

4) How likely would it be that you ask her to erase the recording of the conversation you just had? Not likely 1 2 3 4 5 Very likely

5) How important is it that she asks for your permission to copy the conversation to a tape? Not important 1 2 3 4 5 Very important

6) How important is it that she asks for your permission to play the recorded conversation to someone else? Not important 1 2 3 4 5 Very important

7) Do you consider the conversation you were conducting with her confidential? Not confidential 1 2 3 4 5 Very confidential

8) Your Age Range: 18-29 30's 40's 50's 60 or over

9) Your Sex: M F

10) Your Occupation: ______ 11) Today's date: __ / __ / 04 ☞ turn card

0216 0216

with a consent notice to participate in the study. The proxy asked the participant to fill out the survey immediately if possible, to increase recall accuracy. Otherwise, the survey portion of the card was return addressed on the backside and could be mailed back at the participant's convenience (a postage stamp was affixed for this purpose). The questionnaire included six questions on a 5-point scale, one multiple-choice question and, on the backside, a blank space for optional comments in addition to the lab's address and space for postage. This structure minimized completion time and, in fact, most participants were able to complete the survey immediately.

Results

Of 45 distributed surveys, 41 usable responses were received. Only one person refused to accept the survey at all. Most surveys were completed immediately, and nine were mailed back afterwards. Twenty-four respondents were in IT or research occupations (students, research scientists, university professors, etc.). The remaining respondents ranged across professions, including: teachers, designers, hairdressers, managers, attorneys, and business owners. Respondents spanned all age groups between 18 and over 60, with the distribution biased towards the younger age groups, reflecting the age group of the proxies. Seventeen respondents were female. There was no strong correlation observed between the opinions expressed by participants and type of occupations, age, nor gender.

Both the proxy and the participant were asked to provide a measure of the sensitivity or confidentiality of the conversation they had just conducted. The participant was asked to rate subjective "*confidentiality*" on a five option scale. Proxies indicated whether the conversation was *sensitive* following precise guidelines given by data protection legislation (i.e., financial, health, religious, and some work related topics are sensitive).[2] The difference between confidentiality and sensitivity will be stressed: significant correlation between sensitivity and confidentiality as indicated by the proxies and the participants was not observed.

Participants wanted to be informed that the recording was happening. They wanted to be informed regardless of the sensitivity of the topic of the conversation as classified by the researcher and the place where it happened (public or not). Participants indicated that it was important that the PAL user (the proxy) ask for permission before using PAL in the conversation that had just occurred. Both these variables correlated highly to perceived confidentiality as well. Participants stated it was important that the person using PAL ask permission before he or she copies or replays the audio to others. There were relatively weak correlations of these two variables with confidentiality, suggesting that there is a concern with what happens with the recording, regardless of its perceived confidentiality. The desire to be informed does not necessarily imply that the participants would have likely asked to erase the recording after the fact. In most situations, people would not have asked to erase the conversation. There appears to be a "confidentiality threshold" in this respect: participants would have asked the user to erase the recording primarily in cases of elevated confidentiality. In most cases of conversations of medium or low confidentiality, they would not have likely asked to delete the recording. Finally, participants indicated that a long retention time would not be an issue. The original estimate of appropriate retention time was in the range of 10 minutes–1 hour, but these participants were comfortable with a much longer retention time. Apparently, participants were more concerned with the misuse of the recording (e.g., by replaying to others) than with its mere storage.

CRITICAL FEATURES FOR USE

Informed by findings from the laboratory and diary studies, the multi-month deployment, the exploration of privacy regulations, and the contextualized survey of conversation partners, the critical features of PAL were uncovered. Many of these features fed into the iterative design process of developing the final prototype of PAL on a mobile phone.

Making PAL Useful

Given the rates of 2.5 and 3 incidents per week as reported by the deployment and the diary study, the need for PAL is justified. Analysis of the stated purpose for recovering the audio provided additional information (see Table 1 for a synthesis).

Information minimization requires collecting the minimum amount of personal information needed by the application. Given that 58% of the diary incidents referred to content within 1 hour, a buffer of up to 60 minutes should suffice, with a 15 minute default. EU and U.S. law diverge in this regard, as ECPA does not make any distinction based on stored information retention time. A more conservative way of looking at this issue would be that of understanding the duration of the "social contract," implicit among parties engaged in a conversation, to determine how long a recording can be maintained after the end of such conversation. This measure relates to the relation between distance and place (Citro, Iglen, & Marrett, 2003): how long does it take to move between places with incompatible social contracts? Because PAL could be abused when crossing place boundaries, the recording should be limited to minimize such risks. While valid from a phenomenological standpoint, it was decided to postpone this assessment, given the unsolved issue of gathering reliable contextual data.

Making PAL Ubiquitous

A mobile/wearable solution for PAL was targeted. The intuition was that the mobile phone would likely be with an individual much of the time (at least during working hours, perhaps also at home). Of the participants in the laboratory study who owned a mobile phone, all but one was carrying it upon arrival for the study. Furthermore, in 79% of the diary entries queried, the participant's mobile phone was on her, or within reach. A study was conducted later and it was learned that people's mobile phones are within an arm's proximity to them only 50% of the time (Patel, Kientz, Hayes, Bhat, & Abowd, 2006), however, the mobile phone still represents the only truly personal and ubiquitous device available today.

The results of all of the studies associated with PAL demonstrated the need for and appropriateness of this service to be wearable, as opposed to environmental. The argument can be made that an audio buffering service in the environment might be preferable for a variety of reasons, including power concerns, better audio quality, and the convenience of users not needing to wear a device. Every participant reported, however, that there are times when it would not be possible for the service to be environmental. Every participant who recorded any entries recorded at least one at a public place or outdoors, where environmental solutions would be difficult. Participants also ex-

Table 1. Purpose for recovering audio reported from initial diary study (total 109 entries in diary study)

Purpose category	**Occurrences**
Forgotten previous details (e.g., making a list, retrieving details)	36 (33%)
Replaying for conversation partner (replaying for person who either spoke the audio originally or was present to hear it)	20 (18%)
Interrupted (external activity took focus away from important audio)	18 (17%)
Explicit tape recorder behavior (participant was aware prior to the incident that she wanted to record it)	13 (12%)
Distracted (another concurrent activity took attention)	13 (12%)
Relaying information from one partner to another (replaying for person not present when original audio was recorded)	9 (8%)

pressed control concerns about an environmental version of PAL versus a wearable solution. One participant noted, "[I would] rather have the control of it being on my person."

While advocating a wearable solution, however, participants were not interested in a completely separate device but instead as a "value added feature" to the mobile phone already owned and carried. Although this may seem obvious in retrospect, it implies the fairly strict requirements that PAL must run unattended on the mobile handset, without recharging for at least a day, and it must not interfere with the call functions of the phone. These requirements are met by the currently deployed prototype, resulting in an arguably ubiquitous service.

Making PAL Usable

The final PAL prototype provides asymmetric backward/forward skip features over the recording, with default values of 10 and 5 seconds, respectively. Although most participants of the laboratory study liked these defaults, the values can be adjusted, and anecdotal experience shows that individuals do optimize them. Effective use of fast forward or rewind skimming features were not observed during the laboratory study. Considering the limited capabilities of the handset, it was opted to support earmarks instead. The user can set earmarks and then use the backward/forward skip buttons to traverse these earmarks or simply navigate without using them. One issue identified in the laboratory study related to the mapping of the pair of navigation buttons is: there is no "natural association" between the buttons and backward and forward navigation. The variety of ways the handset can be mounted on a belt or carried in the pocket or purse exacerbates this issue. Thus, the "never-wrong" mapping described in the final prototype was opted for.

Making PAL Socially and Legally Acceptable

The common opinion that people must adapt to technological evolution by changing their social expectations is not necessarily endorsed. However, a case could be made that PAL does not impinge on constitutional rights and that, in the long term, practice could show the harmlessness of this application, granted specific guarantees, namely, small recording radius, short buffer length, and some form of notification to the conversation partners. It is stressed that it is not in the scope of this chapter to provide conclusive legal opinions—a task best left to courts and DPAs. The purpose is to provide a balanced, if necessarily concise, overview of PAL's social and legal impact.

A number of different stakeholders can be identified with regards to PAL. Three are considered: the user, conversation partners, and unrelated third parties. Considering the third category, diary results indicate that 69% of the entries related to recordings in public or semi-public spaces and 44% stated that other, unrelated people were present. These figures support the concern with third-party privacy which contrasts with the fact that the vast majority of participants neither were preoccupied with a third party's privacy nor with that of the conversation partner. These observations are particularly interesting because they diverge from legislation in force. ECPA does prohibit capturing a third party's conversation when the owner of the device is not part of that conversation and the conversation takes place with reasonable expectation that it is not being intercepted (e.g., non-public space). On the other hand, it must be noted that the perceptual properties of sound might not grant constitutional basis (in the U.S.) for an expectation of privacy in public space, as suggested among others by numerous cases adopting the "plain view" rule. This could allow adapting surveillance legislation to permit limited memory aid devices such as PAL.

Interface affordances and information retention policies greatly impact social acceptability. Altering the coverage of the microphone is an essential factor of a proportionality determination, as suggested by analogous DPA opinions involving personal uses of video surveillance (namely, outdoor camera units at home entrances) (Ackerman, Cranor, & Reagle, 1999). Likewise, DPAs have used retention time and deletion policies to evaluate the social impact of surveillance applications. Completely eliminating the risk of recording third parties' conversations is extremely difficult, given the characteristics of sound transmission. PAL arguably presents lower risks than traditional audio recorders; the retention properties of this application do support the claim that PAL does not serve archival purposes, nor does it vastly facilitate surveillance.

In the relationship with conversation partners, informed consent is one fundamental tool of social action, embodied in privacy law. Its implementation presents formidable technical and usability challenges. In this case, anecdotal evidence collected during the deployment suggests that the participants have, over time, renounced to preventively explain or ask permission to use the service. At times, participants turned off the device due to social pressure. Both observations support the previous findings from the diary study. This could hint at a gradual adaptation to the technology and the adoption of appropriate social behavior, similarly to what is currently happening with camera phones.

Although strong feedback was not received from the participants requesting that PAL provide a notification cue while recording, in view of the considerations, it was decided that such function be incorporated in the deployed handsets. When recording, the outer LED integrated in the round ornament on the phone shell (see Figure 1) lights up red. During playback, the light turns green. Although recording is usually associated with a red indicator, there is awareness that people might not understand its meaning and those users could obviously conceal the LED as well as the recording device: the user remains ultimately responsible for abiding to the social contract and mores.

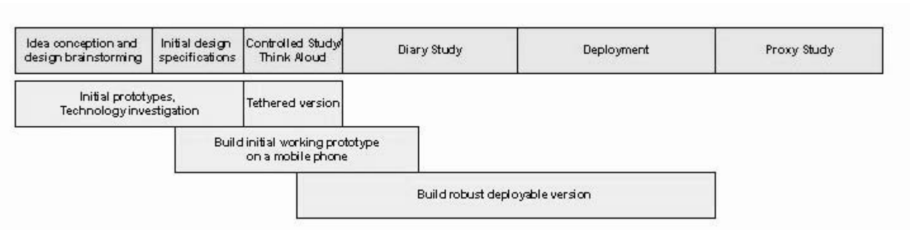
FUTURE TRENDS IN DESIGNING FOR MOBILE APPLICATIONS

As computing moves more off the desktop and mobile devices become more computationally powerful, we as a community must begin to examine the ways in which we design, develop, and evaluate mobile applications. Many lessons can be learned from the 4 year process of designing, developing, and evaluating the Personal Audio Loop. Several important features may be valuable for other researchers and designers to consider. First, the team had a fundamental advantage in conducting this type of design work based on the diversity of skills and perspectives represented. Expert mobile developers used their technical skills to prototype working applications on the mobile phone very quickly. Additionally, they were able to rapidly iterate on the software as the project progressed. Designer experts in conducting formative and qualitative studies as well as those who were knowledgeable in legal and regulatory issues, determined how the design complied with laws and social considerations. This team of varied expertise afforded a mixed-method evaluation of the designs, including the lab studies, diary studies, deployment studies, and the proxy study.

Concurrent development and evaluation of prototype systems throughout the project also contributed significantly to its success. The timeline outlines the general schedule for the phases of the project (see Figure 5). Notably, the development of the working prototype occurred concurrently with the laboratory study and the diary studies. The development of the tethered prototype allowed these studies to be conducted while the development occurred.

Finally, this project included the use of not only tried and true methods but also those that were invented or adjusted for the particular needs of mobile applications, most significantly, the proxy study. Many mobile applications will be used in public settings, and thus exploration of direct situated public reaction was essential. Few well tested techniques would have provided this information. Using proxies in a field study, however, provided the data to allow for a determination of bystander reaction. This study revealed some surprising results (such as bystanders not caring how long data is kept) and was essential to the project. These types of studies, relatively lightweight, could be deployed by several people over a short amount of time with minimal effort.

Figure 5. Timeline of the entire development cycle. The blue cells indicate the studies conducted and the yellow cells show the development phases in parallel



Despite its many advantages, the approach also left room for improvement upon which must also be reflected on. First, there still exists a need for a longer, richer, more in depth deployment study with more users. Several people using PAL as part of this project carried an additional mobile device, and thus it would be difficult for them to adjust to having a new device primarily for this service for the long term. Additionally, it was not deployed for long enough to determine how long it would take for the novelty effect to wear off. Adding more users to the deployment study would also have provided better results. Though choosing a wide range of people was tried, it was difficult to make categories of user types with such a small sample and with such infrequent use.

When considering applications for mobile phones and other mobile devices, designers and developers must recognize the varied approaches and skills that must be applied to a particular project. The Personal Audio Loop was created based on formative and summative evaluations, legal and social considerations, and technical advances and expertise. Without this multi-angled approach, the project would not have been possible. In the future, it is expected that more project teams use similar processes and team compilations to push further the boundaries of usability, usefulness, novelty, and adoption in the mobile applications space.

CONCLUSION

Based on controlled and field studies of use of a mobile audio-based memory aid, it is concluded that not only is the service desirable for users, but also that its implementation on a mobile phone is possible and usable. Users find the information needed in less time than they reported being willing to spend. They need this service at least once a week and they are willing to carry a mobile phone most of the time to have access to it. The analysis shows that this application falls within a legal "gray area" and that its legality cannot be definitively asserted or denied. The interface and retention characteristics of the application, along with observation of initial deployment, suggest that the application might be socially acceptable.

Participants stated that awareness about PAL was important to allow "boundary-setting" to occur. They were not concerned as much by retention time as with potential misuse of the recordings. They also stated that they would have rarely asked to delete a recording after the fact. These observations have broad consequences because they suggest that traditional privacy guidelines and quantitative privacy policies may not be appropriate or sufficient for the development of this kind of personal ubicomp applications and that designers should focus on the purpose of use of information and interpersonal dynamics instead.

REFERENCES

Ackerman, M., Cranor, L., & Reagle, J. (1999). Privacy in e-commerce: Examining user scenarios and privacy preferences. In *Proceedings of the ACM Conference on Electronic Commerce 1999* (pp. 1–8). New York: ACM Press.

Arons, B. (1993). SpeechSkimmer: Interactively skimming recorded speech. In *Proceedings of UIST'93: The 6th Annual Symposium on User Interface Software and Technology* (pp. 187–196). New York: ACM Press.

Citro, C. F., Iglen, D. R., & Marrett, C. B. (Eds.) (2003). *Protecting participants and facilitating social and behavioral sciences research*. Washington, DC: National Academies Press.

Czerwinski, M., Horvitz, E., & Wilhite, S. (2004). A diary study of task switching and interruptions. In *Proceedings of the CHI 2004: ACM Conference on Human Factors in Computing Systems* (pp. 175-182). New York: ACM Press.

Deitz, P., & Yerazunis, W. (2001). Real-time audio buffering for telephone applications. In *Proceedings of the UIST 2001: The 14th Annual ACM Symposium on User Interface Software and Technology* (pp. 193–194). New York: ACM Press.

Goffman, E. (1966). *Behavior in public places.* New York: Free Press.

Hayes, G. R., Patel, S. N., Truong, K. N., Iachello, G., Kientz, J. A., & Farmer, R. (2004). The personal audio loop: Designing a ubiquitous audio-based memory aid. In *Proceedings of the Mobile HCI 2004: The 6th International Symposium on Mobile Human-Computer Interaction* (pp. 168–179). Berlin/Heidelberg: Springer.

Hindus, D., & Schmandt, C. (1992). Ubiquitous audio: Capturing spontaneous collaboration. In *Proceedings of the CSCW'92: The ACM Conference on Computer Supported Cooperative Work* (pp. 210–217). New York: ACM Press.

Iachello, G. (2006). *Privacy and proportionality.* Ph.D. Dissertation, Georgia Institute of Technology, Atlanta, GA.

Iachello, G., Truong, K. N., Abowd, G. D., Hayes, G. R., & Stevens, M. M. (2006). Event-contingent experience sampling to evaluate Ubicomp privacy in the real world. In *Proceedings of the CHI 2006: ACM Conference on Human Factors in Computing Systems* (pp. 1009–1018). New York: ACM Press.

Patel, S. N., Kientz, J. A., Hayes, G. R., Bhat, S., & Abowd, G. D. (2006). Farther than you may think: An empirical investigation of the proximity of users to their mobile phones. In *Proceeedings of the Ubicomp 2006*: *The 8th International Conference on Ubiquitous Computing* (pp. 123-140). Berlin/Heidelberg: Springer.

Scollon, C. N., Kim-Prieto, C., & Diener, E. (2003). Experience sampling: Promises and pitfalls, strengths and weaknesses. *Journal of Happiness Studies, 4*, 5-34.

Terrell, T., & Jacobs, A. (2002). Privacy, technology, and terrorism: Bartnicki, Kyllo, and the normative struggle behind competing claims to solitude and security. *Emory Law Journal*, *51*(4), 1469–1511.

Vemuri, S., Schmandt, C., Bender, W., Tellex, S., & Lassey, B. (2004). An audio-based personal memory aid. In *Proceedings of the Ubicomp 2004: The 6th International Conference on Ubiquitous Computing* (pp. 400-417). Berlin/Heidelberg: Springer.

Wheeler, L., & Reis, H. T. (1991). Self-recording of everyday life events: Origins, types, and uses. *Journal of Personality, 59*(3), 339-355.

KEY TERMS

Audio Buffer: Some quantity of constantly recording audio that can be accessed for a set period of time before being deleted automatically.

Deployment Study: A study that asks people to use a new piece of technology as part of their daily lives and activities for some extended period of time and typically includes interviews, surveys, and other means of measuring user response to the technology.

Diary Study: A study that asks people to keep a diary or journal of their interactions with a computer system, any significant events or problems during their use of a system, or other aspects of their working life.

Event-Contingent Sampling: A variation of the experience sampling method in which users are probed upon the occurrence of a particular event.

Experience Sampling: A set of techniques to capture people's behaviors, thoughts, or feelings as they occur in real-time, also known as a pager study.

Memory Aid: Tool or other support designed to augment human memory or correct for human memory error.

Personal Audio Loop: Mobile memory aid application buffering audio on mobile phone.

Usability: The effectiveness, efficiency, and satisfaction with which specified users of a particular piece of technology can achieve specified goals. Often considered a basic requirement of all new interactive applications and sometimes regulated using ISO9241 11.

Usefulness: As applied to mobile applications, usefulness has to do with the specific utility of that application in the user's life. Often, this level of usefulness directly impacts the adoption of the mobile application or device and may be measured by deploying the technology to such users.

ENDNOTES

1 The United States do not have DPAs specifically appointed to examine privacy issues.

2 Proxies were not asked to rate confidentiality because it was wanted that they collect as objective information as possible.

Chapter LXII
Visualisation of Meeting Records on Mobile Devices

Saturnino Luz
Trinity College Dublin, Ireland

Masood Masoodian
The University of Waikato, New Zealand

ABSTRACT

Multimedia data such as music and video are key forces behind the widespread use of mobile devices today. However, the usefulness of mobile access to multimedia sources extends beyond entertainment. The popularisation of teleconferencing and collaborative technologies on the desktop, for instance, has opened the possibility of recording online meetings for later review. Yet, despite the fact that mobile technology could enable convenient access to recorded meetings, research on interfaces for visualisation and access to meeting recordings on portable devices is currently very scarce. This chapter discusses issues on design, implementation, and evaluation of such interfaces. The discussion is illustrated with a case study and lessons learned from work on the HANMER (Hand-held Meeting Browser) system.

INTRODUCTION

Time-based (or *continuous*) media such as audio and video currently account for a large part of data handled and presented by mobile devices, from PDAs to mobile phones to game consoles. These data are usually of an entertainment nature: most audio data on mobile devices consist of music and audio books, and most video data consist of professionally or home-made movies. In addition to time-based media, there are static media which convey what one could term "space-based" data, which comprise still pictures and text, including whiteboard data, SMS, and text documents on PDAs and mobile phones (Harada, Naaman, Song, Wang, & Paepcke, 2004; Lam & Baudisch, 2005; Bederson, Clamage, Czerwinski, & Robertson, 2004; Masoodian & Budd, 2004; Masoodian & Lane, 2003). While presentation and visualisation of, and access to space-based data can be done in a variety of ways, presentation of time-based data predominantly conforms to a "tape recorder metaphor" whereby playback is essentially sequential and a time-line is used as reference for non-sequential access.

More recently, another kind of recorded time-based data has started to emerge. As teleconferencing and computer-supported co-located meetings become common in the work place, meetings are increasingly being recorded for later review by participants or viewing by those unable to attend them (Chiu, Boreczky, Girgensohn, & Kimber, 2001; Chiu, Kapuskar, Reitmeier, & Wilcox, 1999; Tucker & Whittaker, 2005). Recorded meetings are a special kind of time-based data in that meeting participants often employ artefacts of an essentially spatial and persistent nature, such as agendas, minutes, and private or shared notes, at identifiable points along the time-line. Despite the existence of such "structuring elements," achieving effective presentation and visualisation of recorded meetings is far from trivial on any kind of computing platform. It often does not suffice, for instance, to present a textual summary of outcomes or minutes, no matter how accurate, since in reviewing a meeting one is typically interested in *why* certain decisions were made rather than simply *what* those decisions happened to be (Moran et al., 1997). Many attempts have been made at supporting the activity of "browsing" recorded meetings by indexing their speech content (Stiefelhagen, Yang, & Waibel, 1999; Waibel et al., 2001; Tucker & Whittaker, 2005), temporally aligning text and speech streams (Masoodian & Luz, 2001), or simply by improving sequential access (Arons, 1993). These attempts mainly targeted information retrieval aspects and the resulting systems essentially comprised improved time-line interfaces, or combined basic time-line access with indexed textual information and meta-data. With the possible exception of SpeechSkimmer (Arons, 1993), those interfaces cannot be easily adapted for use on small mobile devices, where screen space constraints are a primary concern.

In this chapter, a novel approach to the design of meeting browsing interfaces for mobile devices is described. This approach explores the interrelation between the time-based component of meetings and its space-based artefacts in order to implement a view of recorded activity which supports structured, non-linear indexing and access to speech and textual elements. The ensuing visualisation components depart from the rigidity of basic time-line interfaces and are particularly suited for presentation on hand-held mobile devices which lack the input and output capabilities of the conventional computers for which almost all other existing meeting browsers have been developed. For simplicity, attention is confined to speech as the recorded time-based modality. However, the techniques and observations made here are expected to apply equally to video. In fact, towards the end of the chapter, the ways in which recorded video could be profitably exploited as a source of space-based data and access points in the general indexing strategy proposed are discussed.

This chapter is organised as follows. The next section describes the context of the application area and reviews the literature on meeting browsing and relevant works in the area of multimedia access on small mobile devices. This is followed by a presentation of the overall framework and paradigm for meeting browsing on mobile devices, a description of the HANMER prototype, along with an outline of its design process. A section on evaluation and a discussion of results follows. Finally, future trends, perspectives, and conclusions are discussed.

BACKGROUND

Handheld technology is already having a big impact on how people interact and work. It is suggested that these mobile devices have "the potential to provoke even more radical changes in work practices and encourage an even greater level of mobile work and distributed collaboration [than the Internet]" (Perry, O'Hara, Sellen, Brown, & Harper, 2001). Therefore, it is not surprising that people are gradually starting to consider meeting using their mobile devices, or at least record and share work-related information using such devices in situations where mobility is essential (Bergqvist, Dahlberg, Ljungberg, & Kristoffersen, 1999). In recent years, various systems have been developed which utilise mobile devices in facilitating sharing of meeting-related information between co-workers (Davis et al., 1999) or bridging the gap between face-to-face and online meetings (Wiberg, 2000; Wiberg, 2001).

All these developments are leading towards an even greater need for recording of meeting data, whether those meetings are held face-to-face, online, or while on the move. The fact that recording and reviewing of meetings are becoming common is attested by the amount of research targeting information retrieval and visualisation of meeting data (Tucker & Whittaker, 2005; Waibel et al., 2001; Morgan et al., 2001; Geyer, Richter, & Abowd, 2003; Wellner, Flynn, & Guillemot, 2004; Erol, Lee, & Hull, 2003; Luz & Roy, 1999; Masoodian & Luz, 2001) and the ongoing interest in speech retrieval interfaces (Arons, 1993; Coden, Brown, & Srinivasan, 2002; Whittaker et al., 2002).

The dominant paradigm in meeting browsing is founded on modality translation. Systems that adhere to that paradigm (Waibel et al., 2001; Wellner et al., 2004; Tucker & Whittaker, 2005) attempt to convert time-based data into suitable space-based representations. Audio is converted into text through automatic speech recognition (ASR) and video is indexed through key frames (Zhang, Low, & Smoliar, 1995; Stiefelhagen et al., 1999; Waibel et al., 2001) often complemented by audio and language processing (Pfeiffer, Lienhart, & Effelsberg, 2001; Kazman, Al-Halimi, Hunt, & Mantey, 1996). Waibel et al. (2001) describe a system that displays meeting transcripts synchronised on a time-line with their corresponding sound or video files, along with a form of textual meeting "summary" consisting of salient fragments of the ASR-transcribed speech. Another system based on ASR is the Ferret Media Browser (Wellner et al., 2004). Ferret's user interface provides ASR transcripts, keyword search, and speech segmented according to speaker identity. Its distinctive feature is the ability to dynamically insert or remove media streams during browsing. The MeetingViewer (Geyer et al., 2003) is a browser for meetings recorded with the TeamSpace (Richter et al., 2001) online conferencing system. TeamSpace records and time-stamps all actions performed by participants during a meeting. Metadata are used by MeetingViewer for indexing the recording. Actions and metadata are displayed on a time-line. COMAP (Masoodian & Luz, 2001) is a system for browsing time-stamped recorded online meetings in which speech and text are used as communication media for a collaborative writing task. The COMAP user interface displays a step-function graph which summarises the participants' speech and editing activities by providing an indication of the intensity of text and speech interactions. The textual outcome of the writing task and the meeting's speech content can both be accessed through this graph.

The meeting browsing systems described share several user interface features. These include: focus on speech and text for browsing and visualisation support, conversion of time-based data to a space-based medium in the form of a time-line representation augmented with textual components (extracted via ASR or action time-stamping, or both), and support for playback of continuous data in parallel with exploration of the space-based medium. In systems that also index video, the emphasis has been on identifying key frames which play a role analogous to text annotation in that they help situate the search on a time-line (Zhang et al., 1995; Kazman et al., 1996). While these features seem suitable for browsing and searching recorded meetings on desktop computers, annotated time-line representations are notoriously space-consuming and therefore unsuitable for small mobile devices.

Unlike multimedia meeting browsing systems, interfaces based on improved audio access, such as SpeechSkimmer (Arons, 1993), are not significantly affected by screen-space constraints. However, it could be argued that by failing to support visualisation and access to static components, such interfaces do not provide adequate support for meeting browsing. Moreover, even for retrieval of recordings consisting entirely of speech, there has been a trend towards using text as an access modality to speech data (Whittaker et al., 2002).

The issue of catering for the requirements of mobile users when designing meeting browser interfaces and the need for systems that can operate on resource-constrained devices has been acknowledged in the literature (Masoodian, Luz, & Weng, 2003; Tucker & Whittaker, 2005). Yet, despite the development of a wide range of meeting recording systems and various kinds of recorded meeting browser and access tools for conventional desktop computers, very little research has been carried out on developing similar systems for mobile devices.

The main focus of related research on systems for mobile technology, with any relevance to the work and meeting browsing in general, has been on issues related to media adaptation. This consists of tailoring multimedia presentation to the constraints imposed by mobile devices, including reduction of image quality or replacement by schematic representations (Rist & Brandmeier, 2002), selection of key-frames for delivery or browsing of video (Liu & Choudary, 2004; Liu, Hua, Zang, Tong, & Lu, 2005; Kamvar, Chiu, Wilcox, Casi, & Lertsithichai, 2004), and selection of "regions of interest" for browsing and access to image collections (Hua et al., 2005; Patel, Marsden, Jones, & Jones, 2004). Although these techniques are useful in the particular domains for which they were designed, no unified paradigm has emerged from them which could encompass access to meeting data on small mobile devices.

VISUALISING MEETING RECORDINGS ON SMALL SCREENS

This section describes a design concept for browsing and retrieval of information from meeting recordings, while addressing the issues discussed. This design concept consists essentially of enabling sharing of screen real-estate between a continuous representation component (as in standard time-lines) and contextualised textual data. This idea is embodied in a Hand-Held Meeting Browser. HANMER thus implements a trade-off between the linearity inherent to audio browsing and the parallelism needed for contextualisation in order to attain efficient use of screen resources. Evaluation shows that this compromise does not impair the user's ability to interpret time- or space-based media.

The main components of the HANMER interface are a combined speech and text *action visualisation* component and a *text selection* component. The action visualisation component presents a compact representation of speech turns and text editing actions performed during the meeting. This component is realised in two basic forms in the interface: a *radar pane* which displays the entire recording and a *zoom pane* which displays details of a selected interval of the recording. The user controls what is displayed on the zoom pane by selecting a viewport on the radar view. The zoom pane, by its turn, enables the user to play back the audio or view the text relevant to a particular speech segment. The text selection component acts as a space-based counterpart to the zoom pane by allowing the user to visualise, on the radar pane, all speech segments related to the selected text segment. The screen space is thus shared between the radar pane, which is always visible on the lower half of the screen and helps situate the browsing activity and (on the upper half of the screen) either the zoom view or the text selection component. The parallel nature of the speech modality can be exploited so that audio playback and text browsing may be performed simultaneously, allowing the user to discover further relationships between the two media streams.

In order to tie together the space-based and the time-based media, one needs extra elements to add parallel structure to the continuous data. One such element is provided by the concept of *temporal neighbourhoods*, which is described in the next section.

Media Neighbourhoods and Content Mapping

In Luz & Masoodian (2005) a general strategy for relating time- and space-based media is proposed. It consists of mapping the contents of one medium to the other at two basic levels: *temporal* and *contextual*. Briefly, temporal mapping adds a time element to the static data through synchronisation and time-stamping, and contextual mapping adds structure to the continuous data through a variety of content analysis techniques such as ASR and natural language processing. The relations thus generated are called *temporal neighbourhoods* and *contextual neighbourhoods*. Due to the resource limitations of mobile devices, the mobile meeting browser described here only employs temporal neighbourhoods.

Temporal neighbourhoods can be defined in a number of ways (Bouamrane, Luz, & Masoodian, 2006; Luz & Masoodian, 2005; Bouamrane, Luz, Masoodian, & King, 2005). The definition adopted for HANMER (Masoodian et al., 2003) is possibly the simplest. Two types of objects are defined to start: *speech segments* and *text segments*. For simplicity, a speech segment can be defined as an individual audio interval delimited by silence. Similarly, a text segment can be defined as a paragraph or chunks of text delimited by two sets of consecutive line breaks. Given a set $T = \{t_1,...,t_{|T|}\}$ of text segments, and a set of speech segments $S = \{s_1,...,s_{|S|}\}$, temporal neighbourhoods are determined by interval overlap as follows:

A *temporal text-to-audio mapping* is a function $tn : T \rightarrow 2^S$ defined so that $tn(t_i) = \{s_j : t_i^s \leq s_j^s \leq t_i^e \text{ or } t_i^s \leq s_j^e \leq t_i^e \text{ or } s_j^s \leq t_i^s \leq s_j^e \text{ or } s_j^s \leq t_i^e \leq s_j^e\}$, where t_i^s and t_i^e denote the start and end time of segment t_i.

In other words, a segment of audio is a *temporal neighbour* of a text segment if that audio segment was recorded while the text segment was being modified or pointed at by the participants. This definition includes all relations between text and audio segments allowed in Allen's time interval algebra (Allen, 1984) except for the *before* relation and its converse. Text segments can also be retrieved using audio as a starting point by simply inverting the mapping or defining an *audio-to-text* retrieval function:

$$tn_a : S \rightarrow 2^T, \text{ so that } tn_a(s_i) = \{t_j : s_i \in tn(t_j)\}.$$

A temporal neighbourhood is the relation $\tau \subseteq S \times T$ induced by *tn*. The visualisation described in the following section is meant as a visual counterpart to the temporal neighbourhood concept. The time-stamping and segmentation which are necessary in order to determine a meeting's temporal neighbourhoods can be automatically obtained in a variety of ways through signal processing and data mining methods, the discussion of which is outside the scope of this chapter—see Luz and Masoodian (2005), Foote (1999), and Coden et al. (2002) for overviews of relevant methods. In order to guarantee accuracy, the data employed in this study have been recorded, time-stamped, and segmented by a collaborative writing tool specially designed and implemented for that purpose (Luz, Bouamrane, & Masoodian, 2006; Masoodian, Luz, Bouamrane, & King, 2005).

As will be shown, temporal neighbourhoods are conducive of a kind of browsing activity which are called *non-linear browsing*. This kind of browsing enables the user to follow links between text and speech segments independently from the constraints imposed by the continuous media.

Design of a Meeting Browser for Mobile Devices

As described, the concept of temporal neighbourhoods was chosen as the basis for visualisation of meeting data, which would provide browsing and access to recorded speech and textual meeting content. An initial prototype meeting browser was then developed for desktop computers with conventional displays. This prototype, called COMAP (Masoodian & Luz, 2001), used a classical time-line visualisation (Tufte, 2001), a technique extensively applied to the display of temporal data, including patient history (Plaisant, Milash, Rose, Widoff, & Shneiderman, 1996) and multimedia streams (Harrison, Owen, & Baecker, 1994). COMAP's goal was to highlight temporal neighbourhood links between text and speech recordings of the meeting contents.

This visualisation also formed the basis of the first HANMER mobile meeting browser (Masoodian et al., 2003). However, the initial use of this prototype demonstrated that its underlying visualisation (see Figure 1) suffered from a major limitation which would make it unsuitable for devices with small screens when the number of meeting participants increased to more than a couple. The basic problem is that this visualisation needs to show as many bars per participant as there are media. Therefore, in a meeting with three participants communicating through text and speech, for instance, the conventional time-line would comprise six horizontal lines. As the number of participants increases this type of representation quickly becomes uninterruptible on a small screen. For the meeting browsing task, it is essential that the user be able to quickly identify interval overlaps and concurrent activity, since it is usually in such intervals that most events of interest take place. On a time-line interface consisting of multiple streams, this ability translates into being able to accurately recognise relative position of vertical bars (vernier task). Research on visual hyperacuity has shown that although the human eye is remarkably good at this task (Westheimer & McKee, 1977), visual sensitivity to spatial alignment decreases as the distance between the targets increases (Waugh & Levi, 1995). In multiple-stream time-lines this problem is further aggravated by the fact that alignment often needs to be recognised across several parallel lines.

In order to minimise this problem, a second prototype was developed which employs a mosaic-style visualisation component (see Figure 2). In this visualisation, the space allocated to a particular time-line is used proportionally to show the contributions of each of the participants in terms of that medium, and therefore, only a single horizontal bar is required per medium. Consider, for instance, how speech turns of a

Figure 1. Standard time-line representation of a two-participant meeting

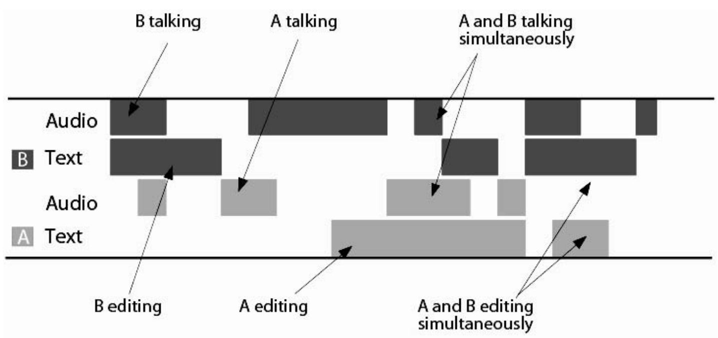
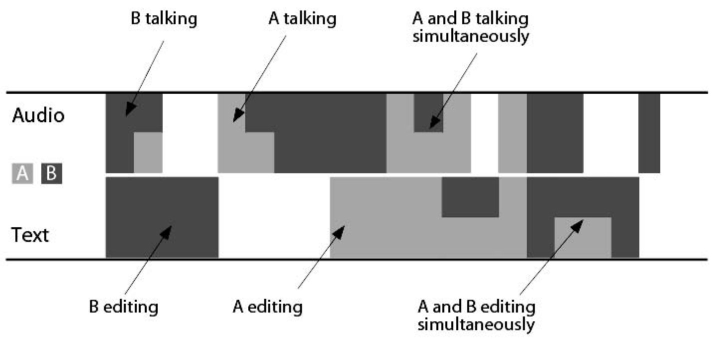


Figure 2. Mosaic-style representation of a two-participant meeting



two-participant dialogue would be represented by this visualisation style: if only one participant (say, person A) is talking for a period of time then the audio time-line corresponding to that time period is covered by a rectangle which has the colour allocated to that participant. If participants A and B are talking simultaneously for a period of time, then the rectangle corresponding to that time period is divided vertically into two and each half is covered with the colours associated with A and B. This allows a larger number of participants to be depicted on the mosaic-style visualisation compared to the standard time-line.

It can be readily seen that the mosaic-style visualisation makes better use of screen space by proving a more compact and arguably more natural representation of silences and text inactivity. Instead of representing "silences" (or inactivity) per participant, as in standard time-lines, HANMER adopts the convention that a participant is silent (inactive) unless depicted as speaking (or writing) on the screen. Only real, collective silence (a somewhat rare event in meetings) or text inactivity are explicitly represented (as non-coloured spaces) on the screen. Furthermore, since overlapping concurrent activities are always displayed adjacently, users can take full advantage of visual hyperacuity to quickly spot events of interest.

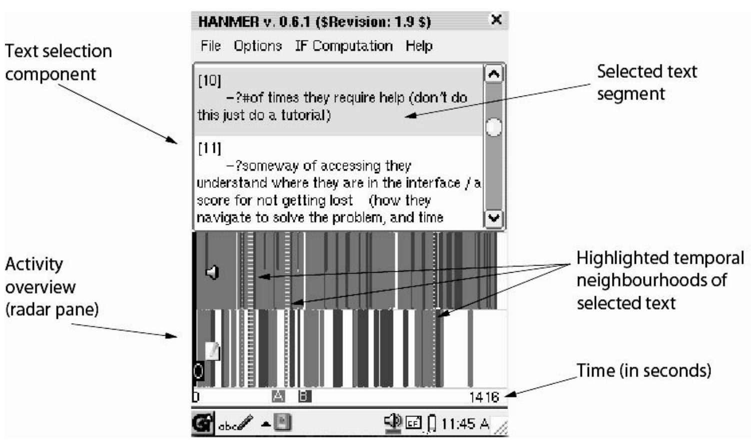
Concurrent activity usually correlates to significance, as decision-making involves discussion and often textual recording of the decision reached. This issue has been addressed elsewhere (Luz, 2002) by proposing a metric which quantifies inter-stream activity. In that respect, another benefit of the mosaic-style visualisation is that it allows users of the system to easily locate parts of the meeting which have higher amount of concurrent activities, as these correspond to the parts which are more colourful and look like mosaics. Furthermore, a quick look at the visualisation can also easily show if any of the participants is more active than the others during the meeting, simply because their colour would be more dominant in the graph.

An Overview plus Detail Interface for Temporal Navigation

The user interface of the HANMER meeting browser is horizontally divided into two regions. The lower half of the screen displays a fixed radar pane, also called the *meeting activity overview*, while the upper half displays *detail* consisting either of a zoomed view of an interval selected on the radar pane or a text selection component.

The initial configuration of the interface shows the text selection component on the upper half of the screen, along with the radar pane of one of the described visualisations (either the classical time-line or mosaic-style visualisation, depending

Figure 3. Text-driven navigation on HANMER



on the user's selection). If the user chooses to use text as the starting point of their browsing activity, selection of a text segment automatically causes its temporal neighbourhoods to be highlighted on the activity overview pane (Figure 3).

Once potentially interesting segments have been located, the user can zoom into specific areas by sliding a selection viewport over the activity overview pane. This causes the text selection pane to be replaced on the screen by the zoom pane (Figure 4) which reacts to user (stylus) input as follows: tapping on the upper half of the zoom pane (audio activity graph) initiates audio playback at the selected point, tapping on the lower half of the zoom pane (text activity graph) causes the zoom pane to be replaced by the text component with the segments which were active at the selected point highlighted.

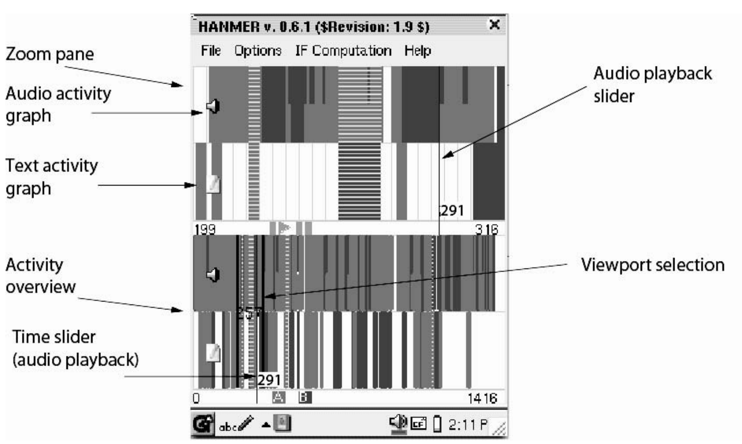
Audio playback is visually indicated on both the zoom and activity overview panes through synchronised sliding vertical lines and persists (unless the user explicitly pauses it) when the text component is visible, as shown in Figure 4. This alternation of audio-text interaction profile and textual content suggests a modality of browsing that differs significantly from the usual linear browsing strategy whereby users would playback the interaction, listening to the speech while observing the effects of keyboard and pointing device interactions on the textual medium. The natural strategy for users of HANMER is to start the search through text, say, by selecting certain topics, and then play, in turns, those speech segments which relate to the selected text segments in order to contextualise the information conveyed by the latter. Since participants often refer back to text segments in speech (e.g., by pointing at them), the user also has to be able to visualise these text referents as needed. The user might start browsing by selecting, for instance, the following collaboratively written text fragment (*t1*), extracted from the corpus of student-supervisor meetings (Luz et al., 2006):

(t1) Hypothesis: Visualisation once understood by user allows the user to do all tasks done by text method just as well but also allows user to make estimates and determine how events interrelate to each other in ways that a text only interface could not.

Segment *t1* is in the temporal neighbourhood of six audio segments. In one such audio segment (*s3*), the speaker makes a deictic reference to text segment *t4*, as shown in the following transcribed speech fragment:

(s3) hmm [pause] where were we? [pause] we were talking about you doing this part [point at t4], which is testing hmm [pause] or finding out user feedback on individual parts then add that to the main bit [...]

Figure 4. Zoom pane and audio navigation on HANMER



Text segment *t4*, by its turn, is linked to another 12 audio segments spread across the entire meeting, thus indicating that *user preferences* was a major topic of discussion. Each of the audio segments in the neighbourhood of *t4* is in the neighbourhood of a number of other text segments, and so on.

(t4) (also find out if general user preference exist as far as a number of interface options i. do they like clocks on turning points or on the line; [...]

Such linking patterns between text and audio, as demonstrated by the example, can be represented by tree structures which can evolve along several dimensions, beyond what is supported by existing linear meeting browsers. It is interesting to note that, in addition to speech links, this technique reveals natural clusters of text segments which are not necessarily in sequence in the original text, as is the case of *t1* and *t4* in the example. Although this inter-connectivity of speech and text can be exploited beneficially by meeting content browser and retrieval systems in general, it is particularly useful for meeting browsing on small mobile devices. Therefore, HANMER is unique in that its visualisation and access functionality are based on the existence of this type of inter-connected meeting content structure.

Although overview plus detail interfaces have been usefully exploited for mobile devices (Bederson et al., 2004), some studies indicate that overview plus detail often leads to worse performance than zoomable interfaces on spatial browsing tasks (Hornbæk, Bederson, & Plaisant, 2002). This is especially true of PDA interfaces (Büring, Gerken, & Reiterer, 2006) where screen real state is greatly restricted. HANMER's overview plus detail interface, however, differs from other such interfaces in that its overview component is the main axis of navigation through which orthogonal space- and time-based media can be accessed and contextualised. Without the overview component, zooming in HANMER would be meaningless, or at least extremely confusing. The question is, then, whether users of HANMER find its visualisation style and non-linear browsing paradigm effective for accessing recorded meeting content. This question is investigated in the next section.

EVALUATION

A usability study of HANMER was conducted to evaluate the effectiveness of its underlying visualisation, as well as the mechanisms it provides for accessing recorded meeting information, both in text and audio. In this section, the details of this user study are discussed and its findings are reported on.

Very little literature exists on the evaluation of meeting browsers in general, and to the best knowledge there has been no published research on evaluation of mobile meeting browsers. Most studies of meeting browsers have relied on informal user feedback (Luz & Masoodian, 2004; Erol et al., 2003). More recently, a general *meeting browser evaluation test* (BET) has been proposed in which meeting browsing is defined as the task of attempting to "find a maximum number of observations of interest in a minimum amount of time" (Wellner, Flynn, Tucker, & Whittaker, 2005). Although that proposal provides a useful starting point to systematic evaluation of meeting browsers, it only addresses the issue from an information retrieval perspective, overlooking user interface design issues which are essential for evaluation of mobile technology. The methodology that will be described complements BET by addressing the issues of usability and visualisation of general information (as opposed to specific "observations of interest") about meeting recordings. Therefore, from this perspective, the meeting browsing task can be stated as the activity of visualising recordings of meetings and finding information of interest in such recordings.

Methodology

This study involved 12 participants, 9 of whom were postgraduate students (75%), 2 were academics (17%) and 1 other (8%). Eight of the participants were male (67%), and four were female (33%). The study involved the use of a PDA, and although 7 of the 12 participants (58%) had used a PDA previously, their use of it was not regular (less than once a month). Five of the participants (42%) had never used a PDA before. Interestingly, as the results will demonstrate, this lack of PDA use did not seem to have any effect on the subjects' perception of HANMER or their ability in performing the required task.

At the beginning of each session the participants were given a simple introduction to HANMER using a PDA and they viewed a sample recorded meeting of around 10 minutes in duration. After the introduction, the participants completed a simple background questionnaire, which was then followed by performing the user study task. The sessions took around 30 minutes each.

The user study task consisted of completing a task questionnaire by viewing and accessing the contents of a recorded meeting using HANMER on a PDA. The task questionnaire consisted of 13 questions, the summary of which is shown in Table 1. The recorded meeting used for this study was around 25 minutes in duration, consisting of recorded speech audio (duration of the entire meeting) and a text document of 32 short paragraphs (about three A4 pages in length). The recorded meeting was between a graduate student and his supervisor, discussing the student's experiment design document which was partly written before the meeting.

As mentioned earlier, the aim of this user study was to evaluate the effectiveness of the underlying visualisation, as well as the various functionalities HANMER provides for allowing better retrieval of recorded meeting contents. Therefore, the task questionnaire consisted of different types of questions to allow these various aspects of the system to be evaluated. Questions 1–3 evaluate the radar view of the HANMER visualisation, Questions 4–7 evaluate the paragraph selection and referencing, Questions 8–10 evaluate effectiveness of time-line referencing in relation to speech and text activities, and finally, questions 11–13 focus on information retrieval using HANMER.

Each task question also asked for the users' subjective view of how easy or difficult it was to perform that particular tasks (i.e., find the answer to the question using the system), using a seven-point rating scale with 1 being easy and 7 being difficult.

Results

A possible way of evaluating the effectiveness of HANMER, in terms of its visualisations and functionality, is to measure the accuracy with which the participants managed to respond to various task questions. Table 2 shows the accuracy of responses for each task question, by all the participants. All participants managed to get the correct responses to 9 of the 13 task questions (69%), while question 13 was answered incorrectly only once, and question 10 twice. Considering the fact that 42% of the participants had never used a PDA, while the rest (58%) did not use one regularly, HANMER seems to be, in general, effective in terms of its visualisation and functionality.

One obvious point of concern is clearly the accuracy with which the participants responded to questions 4 and 6. However, both of these questions relied on the ability to see temporal neighbourhoods of a paragraph (i.e., points in time when it had been edited, referenced, etc.), once it was selected, on the time-line radar pane. The current version of HANMER has been implemented in Java for PDAs. However, as the version of Java used in the implementation of the prototype (Personal Java 3.1) does not support transparent layers, a horizontal-line fill box has been used to show selected temporal neighbourhoods in the radar pane (see Figure 3).

Unfortunately this representation of selection is not particularly easy to see on a small display. There is awareness of this problem and a solution to it will be proposed later in this section.

Table 3 shows the performance of each of the participants in terms of the accuracy of their responses to the task questionnaire in total. Once again, it is interesting to note that despite the participants' lack of experience in using PDAs, five of the participants (42%) answered all 13 questions correctly, while three people (25%) answered 12, and a further three answered 11 questions correctly. Only one person answered 3 questions incorrectly. A more careful look at the participants' incorrect responses showed that in most cases, their answers were very close to the correct answers.

Table 1. Summary of the task questionnaire

Q1
Which of the two participants spoke more during the meeting?

Q2
Which of the two participants wrote more during the meeting?

Q3
Did the participants do more writing or talking during the meeting?

Q4
How many times was paragraph 3 referenced (e.g., edited, discussed, etc.) during the meeting?

Q5
Which of the paragraphs 3 or 4 was referenced (e.g., edited, discussed, etc.) more during the meeting?

Q6
At what point in time was paragraph 3 first referenced (e.g., edited, discussed, etc.) during the meeting?

Q7
Which of the two participants was talking when paragraph 3 was first referenced (e.g., edited, discussed, etc.) during the meeting?

Q8
Which of the two participants was talking between 510 and 530 seconds?

Q9
Which of the two participants was writing between 640 and 660 seconds?

Q10
Which paragraph was referenced (e.g., edited, discussed, etc.) between 720 and 730 seconds?

Q11
Fill in the following transcription of the speech spoken by Participant B (Blue) between 985 and 995 seconds.

"Some tasks they can do in the................way but they couldn't do in................way easily, or at all event."

Q12
What is the problem with "helping people" as described by Participant B (Blue) between 230 and 235 seconds?

Q13
Describe the "back button" example given by Participant A (Orange) when paragraph 11 is discussed at some point during the meeting.

Table 4 shows the average subjective ratings for the ease with which participants answered each of the questions in the questionnaire. Ratings were chosen from a scale varying from 1 (easy) to 7 (difficult.) Table 5, on the other hand, shows the percentage of subjective ratings for each scale-point per each question (nb. individual percentages have been rounded down).

These subjective ratings show that most participants found the task questions generally easy to moderately easy, with 10 of the 13 questions (77%) being rated less than 3.5 on the scale. Question 6 was rated the most difficult to answer, partly because of the temporal neighbourhood selection which has already been discussed, and partly because it required finding an exact time in seconds,

Table 3. Accuracy of responses for each participant

Questions	Total (out of 13)	Percentage
P1	13	100
P2	13	100
P3	12	92
P4	12	92
P5	13	100
P6	11	85
P7	13	100
P8	13	100
P9	12	92
P10	11	85
P11	10	77
P12	11	85

Table 4. Participants' average subjective ratings for each question

Questions	Average Rating
Q1	1.25
Q2	2.75
Q3	1.50
Q4	3.33
Q5	1.83
Q6	4.08
Q7	2.33
Q8	2.50
Q9	2.25
Q10	2.58
Q11	3.75
Q12	3.08
Q13	3.83

Table 2. Accuracy of responses for each task question

Questions	Total (out of 12)	Percentage
Q1	12	100
Q2	12	100
Q3	12	100
Q4	7	58
Q5	12	100
Q6	8	67
Q7	12	100
Q8	12	100
Q9	12	100
Q10	10	83
Q11	12	100
Q12	12	100
Q13	11	92

Table 5. Percentage of the participants' ratings for each scale-point

Questions	Scale-points (%)						
	1	**2**	**3**	**4**	**5**	**6**	**7**
Q1	83	8	8	0	0	0	0
Q2	33	25	17	0	17	0	8
Q3	58	33	8	0	0	0	0
Q4	17	25	17	25	0	0	17
Q5	58	25	0	8	8	0	0
Q6	0	25	25	8	17	8	17
Q7	42	33	8	0	8	0	8
Q8	42	25	17	0	0	8	8
Q9	25	50	8	8	8	0	0
Q10	25	17	42	8	8	0	0
Q11	17	17	25	8	0	17	17
Q12	17	25	25	8	17	8	0
Q13	17	8	25	17	8	8	17

on the time-line visualisation. This problem and its solution will be discussed in the following section. Question 4 was interesting in that despite sharing the same problem with question 6, and receiving the highest number of incorrect responses, it was not rated as difficult as question 6, because most incorrect answers were close enough to the correct one, and the participants only needed to count the number of times a paragraph had been active (e.g., edited, referenced, etc.) rather than finding the exact time of those text activities.

Questions 11 and 13 were also rated as reasonably difficult, mainly because most of the participants were not native speakers of English and the quality of the recorded meeting audio was not very high. Therefore, although most participants found the correct answers to questions 11 and 13, a few of them found it a bit difficult to understand what was being said in the recorded audio meeting.

Discussion of the Usability Evaluation

The results of the study are certainly encouraging as they seem to suggest the positive value and effectiveness of HANMER as a tool for accessing recorded meeting information on PDAs. The system certainly has an efficiency value, demonstrated by the fact the study participants completed the 13 task questions in around 10-15 minutes on average—without it, the participants would have had to listen to the entire meeting's audio recording (25 minutes in duration) several times to be able to answer the task questions.

The study also showed a positive user experience, measured through the ease with which the tasks were completed—the subjects found almost all of the task questions easy to moderately easy to complete. Indeed, it is expected that in real-world situations where the users of HANMER would be accessing recorded meetings which they have attended themselves, or at least will know something about them (e.g., in terms of agenda, structure, etc.), the system will be much easier to use and certainly more efficient in providing people with the necessary functionality for finding meeting information without having to listen to the entire audio recording of the meetings or searching through the completed meeting documents.

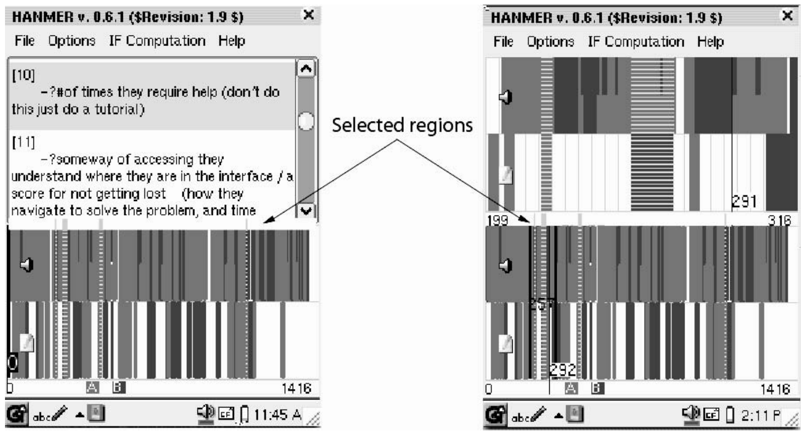
Furthermore, the usability evaluation has not shown any major issues which are not easy to solve. The first problem which is related to how selected temporal neighbourhood of paragraphs are shown on the time-line radar pane can be solved, either by using full transparency when it is implemented by Java Virtual Machines available for PDAs or simply by modifying the interface design of HANMER. A solution would be, for instance, to use the middle area between the radar pane and the text/zoom pane, which is currently occupied by the stop/start and pause buttons of the audio player (see Figure 4), for marking the selected regions of the time-line, as shown in Figure 5. The stop/start and pause buttons of the audio player can simply be added to a context-sensitive menu which would pop up when the stylus is held on, over a point on the audio activity graph.

The second problem which is related to finding exact time references on the visualisation can be solved by showing a time-box that could appear when the stylus is held over a time-line position or as part of a context-sensitive menu when it is shown (usually done when the stylus is held over a point for a longer time than a simple click).

FUTURE TRENDS

The trend towards the lessening of time and distance barriers by emerging mobile technologies points to increasing opportunities and needs for recording, sharing, and accessing meeting information and data. Further research on design and development of mobile meeting technology can foster such opportunities and address some of these needs. Although existing hand-held and mobile technologies are limited in many ways, they already play a key role in supporting human-to-human interaction and collaboration. As mobile devices become more capable in terms of processing power, input, output, and wireless connectivity, they will become even more important as collaborative work technology.

Figure 5. A modified view of selected temporal neighbourhoods of a paragraph



It is expected that meeting browsers such as the ones discussed in this chapter, which combine rich multimedia streams with modern information visualisation, data mining, and language processing techniques, will be crucial tools in future work environments. Since mobility will be one of the defining characteristics of such environments, meeting browsers need to cater for the needs of mobile users. The techniques presented in this chapter represent an attempt to tackle some of the challenges posed by these needs. Although the system described here is currently limited to accessing recorded audio and textual meeting data locally stored on a PDA, it demonstrates the effectiveness of this type of meeting support systems. There are several ways of extending the capabilities of the prototype, or developing other systems, to more fully support mobile meeting reviewing. Some of these are briefly discussed in the remainder of this section.

An important consideration for meeting recording and browsing systems is support for multimedia data, including video, still images, animation, and so forth. Current mobile technology is limited in its support for storage, retrieval, and viewing of large volumes of multimedia data. Although future advances in hardware will solve some of these problems, further research is needed on better recording, indexing, search, and browsing systems dealing with different types of multimedia meeting data on mobile devices. The preceding sections provide references to several research directions in addressing this situation. Few of those, however, are tailored for data originating from recorded meetings which, as has been seen, tend to have a clear underlying structure.

The need for recording and accessing multimedia data becomes more important as we move beyond conventional meetings (i.e., people meeting face-to-face or on-line to work) and start thinking about supporting other forms of group interaction using mobile devices. A system such as HANMER could easily be extended to allow retrieval of recorded lecture presentations using mobile devices, similar to existing systems for desktop computers (Abowd et al., 1996; Apperley et al., 2002; Hürst, 2003; Brotherton & Abowd, 2004). Like conventional meetings, these types of multimedia presentations have an underlying link structure between the presenter's speech, content of individual slides, annotations made on those slides during the presentation, captured whiteboard data, and so forth.

Another area requiring further research and development is that of remote access to recorded meeting data. Since the storage capabilities of mobile devices are still limited, remote access to streamed multimedia data is a particularly important area. In the case of video, there are several issues to be addressed. Firstly, there is the problem of efficient delivery of video over the wireless network—which is beyond the scope of this chapter and is being improved gradually. The second issue is related to retrieval of required segments of video from a mobile device. It is believed that the ideas developed in this chapter are relevant to this aspect of access to video data. Video can be decomposed into separate media streams consisting of various levels of space- and time-based data. Since key frames can be identified through shot boundary detection and hierarchical clusters of colour histograms (Aigrain, Zhang, & Petkovic, 1996), visual indexing can, in principle, be mapped into temporal neighbourhoods the same way as text. Furthermore, image processing techniques can be employed to extract textual information from time-stamped still images, thus providing another indexing source of space-based data which is independent of direct capture through purpose-built collaborative writing tools. The fact that such natural extensions can be easily accommodated within the browsing paradigm proposed, attests to its generality and adequacy.

As a final note, it is also interesting to consider recent developments in portable and mobile devices which have become an integral part of life in their many diverse forms. As yet, it is not clear whether, over time, these devices will merge into some form of "super mobile device" containing all the functionality of the combined devices or whether each of the devices will evolve within their own range. However, one can speculate that as these mobile devices advance and acquire more processing power and on-board memory, users will expect more advanced functionality from them. Therefore, one can imagine that in the future people may wish to load their meeting data onto mobile phones or portable MP3 players, just as they do with their photos, music, and video-clips.

CONCLUSION

This chapter sought to help bridge a gap between existing research efforts on meeting browsing and mobile technology. Its motivation stemmed from a perceived potential for the use of mobile devices in the context of evolving collaborative technologies centred on remote collaboration and mobility, as well as an acknowledged need for techniques to enable users to review meeting recordings while on the move. These issues were addressed by proposing a general paradigm for meeting browsing which satisfies the core information access requirements of the task within the constraints imposed by mobile technology. The proposed paradigm was illustrated through the description of a hand-held meeting browser and its evaluation.

The techniques presented here focused mainly on text and speech as the main sources of meeting data. However, the paradigm proposed is general enough to accommodate most types of data exchanged during collaborative meetings.

REFERENCES

Abowd, G. D., Atkeson, C. G., Feinstein, A., Hmelo, C., Kooper, R., Long, S., Sawhney, N., & Tani, M. (1996). Teaching and learning as multimedia authoring: the classroom 2000 project. In *Proceedings of the Fourth ACM international conference on multimedia (MULTIMEDIA '96)* (pp. 187-198). New York: ACM Press.

Aigrain, P., Zhang, H., & Petkovic, D. (1996). Content-based representation and retrieval of visual media: A state-of-the-art review. *Multimedia Tools and Applications*, *3*, 179-202.

Allen, J. F. (1984, July). Towards a general theory of action and time. *Artificial Intelligence*, *23*, 225-255.

Apperley, M., Jansen, S., Jeffries, A., Masoodian, M., McLeod, L., Paine, L., et al. (2002). Lecture capture using large interactive display systems. In *International Conference on Computers in Education (ICCE'02)* (pp. 143-147). Los Alamitos, CA: IEEE Computer Society.

Arons, B. (1993). SpeechSkimmer: Interactively skimming recorded speech. In *Proceedings of UIST'93: ACM symposium on user interface software technology* (pp. 187-196). Atlanta.

Bederson, B. B., Clamage, A., Czerwinski, M. P., & Robertson, G. G. (2004). Datelens: A fisheye calendar interface for pdas. *ACM Transactions on Computer-Human Interaction*, *11*(1), 90–119.

Bergqvist, J., Dahlberg, P., Ljungberg, F., & Kristoffersen, S. (1999). Moving out of the meeting room: Exploring support for mobile meetings. In *Proceedings of the 6th European Conference on Computer Supported Cooperative Work* (pp. 81-98). Norwell, MA, USA: Kluwer Academic Publishers.

Bouamrane, M.-M., Luz, S., & Masoodian, M. (2006). History based visual mining of semi-structured audio and text. In *Proceedings of the 12th International Multi-media Modelling Conference (MMM'06)* (pp. 360-363). Beijing, China: IEEE Press.

Bouamrane, M.-M., Luz, S., Masoodian, M., & King, D. (2005). Supporting remote collaboration through structured activity logging. In G. C. F. Hai Zhuge (Ed.), *The grid and cooperative computing GCC 2005* (Vol. 3795, pp. 1096-1107). Beijing: Springer-Verlag.

Brotherton, J. A., & Abowd, G. D. (2004). Lessons learned from eclass: Assessing automated capture and access in the classroom. *ACM Transactions on Computer-Human Interaction*, *11*(2), 121–155.

Büring, T., Gerken, J., & Reiterer, H. (2006). Usability of overview-supported zooming on small screens with regard to individual differences in spatial ability. In *Proceedings of the Working Conference on Advanced Visual Interfaces (Avi '06)* (pp. 233-240). New York: ACM Press.

Chiu, P., Boreczky, J., Girgensohn, A., & Kimber, D. (2001). LiteMinutes: an Internet-based system for multimedia meeting minutes. In *Proceedings of the 10th International Conference on World Wide Web (WWW'01)* (pp. 140-149). New York: ACM Press.

Chiu, P., Kapuskar, A., Reitmeier, S., & Wilcox, L. (1999). NoteLook: Taking notes in meetings with digital video and ink. In *Proceedings of the 7th ACM International Conference on Multimedia (part 1)(MULTIMEDIA'99)* (pp. 149-158). New York: ACM Press.

Coden, A. R., Brown, E., & Srinivasan, S. (2002). Information retrieval techniques for speech applications. In *ACM SIGIR 2001 Workshop* (Vol. 36, pp. 10-13). New York: ACM Press.

Davis, R. C., Landay, J. A., Chen, V., Huang, J., Lee, R. B., Li, F. C., Lin, J., Morrey, C. B., Schleimer, B., Price, M. N., & Schilite, B. N. (1999). NotePals: Lightweight note sharing by the group, for the group. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'99)* (pp. 338-345). New York: ACM Press.

Erol, B., Lee, D.-S., & Hull, J. J. (2003). Multimodal summarization of meeting recordings. In *Proceedings of International Conference on Multimedia and Expo, ICME'03* (Vol. 3, pp. 25-28). Menlo Park, CA, US:

Foote, J. (1999). An overview of audio information retrieval. *Multimedia Systems, 7*(1), 2–10.

Geyer, W., Richter, H., & Abowd, G. D. (2003). Making multimedia meeting records more meaningful. In *Proceedings of International Conference on Multimedia and Expo, ICME'03* (Vol. 2, pp. 669-672).

Harada, S., Naaman, M., Song, Y. J., Wang, Q., & Paepcke, A. (2004). Lost in memories: Interacting with photo collections on PDAs. In *Proceedings of the 4th ACM/IEEE-CS Joint Conference on Digital Libraries (JCDL'04)* (pp. 325-333). New York: ACM Press.

Harrison, B. L., Owen, R., & Baecker, R. M. (1994). Timelines: An interactive system for the collection and visualization of temporal data. In *Proceedings of Graphics Interface '94* (pp. 141-148). Banff, Alberta, Canada.

Hornbæk, K., Bederson, B. B., & Plaisant, C. (2002). Navigation patterns and usability of zoomable user interfaces with and without an overview.

ACM Transactions on Computer-Human Interaction, *9*(4), 362–389.

Hua, Z., Wang, X.-J., Xie, X., Liu, Q., Lu, H., & Ma, W.-Y. (2005). AIRE: an ambient interactive and responsive environment for mobile image management. In *Proceedings of the 7th International Conference on Human Computer Interaction with Mobile Devices & Services (MobileHCI'05)* (pp. 343-344). New York: ACM Press.

Hürst, W. (2003). Indexing, searching, and skimming of multimedia documents containing recorded lectures and live presentations. In *Proceedings of the Eleventh ACM International Conference on Multimedia (MULTIMEDIA'03)* (pp. 450-451). New York: ACM Press.

Kamvar, M., Chiu, P., Wilcox, L., Casi, S., & Lertsithichai, S. (2004). Minimedia surfer: Browsing video segments on small displays. In *CHI '04: Extended Abstracts on Human Factors in Computing Systems* (pp. 1371-1374). New York: ACM Press.

Kazman, R., Al-Halimi, R., Hunt, W., & Mantey, M. (1996). Four paradigms for indexing video conferences. *IEEE Multimedia* (1070–986X), 63-73.

Lam, H., & Baudisch, P. (2005). Summary thumbnails: readable overviews for small screen Web browsers. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'05)* (pp. 681-690). New York: ACM Press.

Liu, Q., Hua, Z., Zang, C., Tong, X., & Lu, H. (2005). Providing on-demand sports video to mobile devices. In *Proceedings of the 13th Annual ACM International Conference on Multimedia (MULTIMEDIA'05)* (pp. 347-350). New York: ACM Press.

Liu, T., & Choudary, C. (2004). Real-time content analysis and adaptive transmission of lecture videos for mobile applications. In *Proceedings of the 12th annual ACM International Conference on Multimedia (MULTIMEDIA'04)* (pp. 400-403). New York: ACM Press.

Luz, S. (2002). Interleave factor and multimedia information visualisation. In H. Sharp, P. Chalk, J. LePeuple, & J. Rosbottom (Eds.), *Proceedings of Human Computer Interaction 2002* (Vol. 2, pp. 142-146). London:

Luz, S., Bouamrane, M.-M., & Masoodian, M. (2006). Gathering a corpus of multimodal computer-mediated meetings with focus on text and audio interaction. In N. Calzolari (Ed.), *Proceedings of the International Conference on Language Resources and Evaluation LREC 2006*, (pp. 407-412). Genoa, Italy.

Luz, S., & Masoodian, M. (2004). A mobile system for non-linear access to time-based data. In *Proceedings of Advanced Visual Interfaces (AVI'04)* (pp. 454-457). ACM Press.

Luz, S., & Masoodian, M. (2005). A model for meeting content storage and retrieval. In Y.-P. P. Chen (Ed.), *Proceedings of the 11th International Conference on Multi-media Modeling (MMM 2005)* (pp. 392-398). Melbourne, Australia: IEEE Computer Society.

Luz, S., & Roy, D. M. (1999). Meeting browser: A system for visualising and accessing audio in multicast meetings. In *Proceedings of the International Workshop on Multimedia Signal Processing* (pp. 489-494). IEEE Signal Processing Society.

Masoodian, M., & Budd, D. (2004). Visualization of travel itinerary information on PDAs. In *Proceedings of the 5th Conference on Australasian User Interface (AUIC'04)* (pp. 65-71). Darlinghurst, Australia: Australian Computer Society, Inc.

Masoodian, M., & Lane, N. (2003). An empirical study of textual and graphical travel itinerary visualization using mobile phones. In *Proceedings of the Fourth Australasian User Interface Conference on User Interfaces 2003 (AUIC'03)* (pp. 11-18). Darlinghurst, Australia: Australian Computer Society, Inc.

Masoodian, M., & Luz, S. (2001). COMAP: A content mapper for audio-mediated collaborative writing. In M. J. Smith, G. Savendy, D. Harris, & R. J. Koubek (Eds.), *Usability evaluation and interface design* (Vol. 1, pp. 208-212). New Orleans, LA: Lawrence Erlbaum.

Masoodian, M., Luz, S., Bouamrane, M.-M., & King, D. (2005). RECOLED: A group-aware collaborative text editor for capturing document history. In P. Isaías & M. B. Nunes (Eds.), *Proceedings of WWW/Internet 2005* (Vol. 1, pp. 323-330). Lisbon.

Masoodian, M., Luz, S., & Weng, C. (2003). HANMER: A mobile tool for browsing recorded collaborative meeting contents. In E. Kemp, C. Philip, & W. Wong (Eds.), *Proceedings of CHI-NZ'03* (pp. 87-92). Dunedin, New Zealand: ACM Press.

Moran, T. P., Palen, L., Harrison, S., Chiu, P., Kimber, D., Minneman, S., van Melle, W., & Zellweger, P. (1997). "I'll get that off the audio:" A case study of salvaging multimedia meeting records. In *Proceedings of ACM CHI 97 Conference on Human Factors in Computing Systems* (Vol. 1, pp. 202-209).

Morgan, N., Baron, D., Edwards, J., Ellis, D., Gelbart, D., Janin, A., Pfau, T., Shriberg, E., & Stolcke, A. (2001). The meeting project at ICSI. In *Proceedings of Human Language Technologies Conference*. San Diego.

Patel, D., Marsden, G., Jones, S., & Jones, M. (2004). An evaluation of techniques for browsing photograph collections on small displays. In S. A. Brewster & M. D. Dunlop (Eds.), *Proceedings of Mobile Human-computer Interaction—Mobile HCI 2004, 6th International Symposium,* (Vol. 3160, pp. 132-143). Springer.

Perry, M., O'Hara, K., Sellen, A., Brown, B., & Harper, R. (2001). Dealing with mobility: Understanding access anytime, anywhere. *ACM Transactions on Computer-Human Interaction, 8*(4), 323–347.

Pfeiffer, S., Lienhart, R., & Effelsberg, W. (2001, September). Scene determination based on video and audio features. *Multimedia Tools and Applications*, *15*, 59-81.

Plaisant, C., Milash, B., Rose, A., Widoff, S., & Shneiderman, B. (1996). Lifelines: Visualizing personal histories. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (Chi '96)* (pp. 221-227). New York: ACM Press.

Richter, H. A., Abowd, G. D., Geyer, W., Fuchs, L., Daijavad, S., & Poltrock, S. E. (2001). Integrating meeting capture within a collaborative team environment. In *Proceedings of the 3rd International Conference on Ubiquitous Computing UbiComp'01* (pp. 123-138). London: Springer-Verlag.

Rist, T., & Brandmeier, P. (2002). Customising graphics for tiny displays of mobile devices. *Personal and Ubiquitous Computing, 6*(4), 260-268.

Stiefelhagen, R., Yang, J., & Waibel, A. (1999). Modeling focus of attention for meeting indexing. In *Proceedings of the 7th ACM International Conference on Multimedia (part 1) (MULTIMEDIA'99)* (pp. 3-10). New York: ACM Press.

Tucker, S., & Whittaker, S. (2005). Accessing multimodal meeting data: Systems, problems and possibilities. In *Proceedings of MLMI'04: Machine learning for multimodal interaction* (pp. 1-11). Springer-Verlag GmbH.

Tufte, E. R. (2001). *The visual display of quantitative information* (2nd ed.). Graphics Press.

Waibel, A., Brett, M., Metze, F., Ries, K., Schaaf, T., Schultz, T., Soltau, H., Yu, H., & Zechner, K. (2001). Advances in automatic meeting record creation and access. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing* (Vol. 1, pp. 597-600). IEEE Press.

Waugh, S. J., & Levi, D. M. (1995, October). Spatial alignment across gaps: contributions of orientation and spatial scale. *Journal of the Optical Society of America: Optics, Image Science, Vision*, *12*(10), 2305–2317.

Wellner, P., Flynn, M., & Guillemot, M. (2004). Browsing recorded meetings with ferret. In S. Bengio & H. Bourlard (Eds.), *Proceedings of machine learning for multimodal interaction: First International Workshop (MLMI 2004)* (Vol. 3361, p. 12-21). Martigny, Switzerland: Springer-Verlag GmbH.

Wellner, P., Flynn, M., Tucker, S., & Whittaker, S. (2005). A meeting browser evaluation test. In *CHI'05: Extended abstracts on human factors in computing systems* (pp. 2021-2024). New York: ACM Press.

Westheimer, G., & McKee, S. P. (1977). Spatial configurations for visual hyperacuity. *Vision Research*, *17*(8), 941–947.

Whittaker, S., Hirschberg, J., Amento, B., Stark, L., Bacchiani, M., Isenhour, P., Stead, L., Zamchick, G., & Rosenberg, A. (2002). SCANMail: a voicemail interface that makes speech browsable, readable and searchable. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, *CHI'02* (pp. 275-282). New York: ACM Press.

Wiberg, M. (2000). Bridging physical and virtual group meetings with a PC and multiple hand-held devices. In *Proceedings of CHI'00: Extended abstracts on human factors in computing systems* (pp. 357-358). New York: ACM Press.

Wiberg, M. (2001). RoamWare: An integrated architecture for seamless interaction in between mobile meetings. In *Proceedings of the 2001 International ACM SIGGROUP Conference on Supporting Group Work (GROUP'01)* (pp. 288-297). New York: ACM Press.

Zhang, H., Low, C., & Smoliar, S. (1995). Video parsing and browsing using compressed data. *Multimedia Tools and Applications*, *1*, 89-111.

KEY TERMS

Content Mapping: A function that establishes relations between space- and time-based media. Only temporal mapping is used in this chapter. A temporal text-to-audio mapping relates each text segment to a set of speech segments.

Hand-Held Meeting Browser: A system for browsing meeting recordings on small mobile devices.

Meeting Browsing: The activity of visualising multimedia meeting recordings and finding information of interest in such recordings.

Multimedia Meeting Recording: A digital recording of a meeting consisting minimally of a time-based data stream, typically speech, and a space-based data stream, typically text.

Non-Linear Browsing: A modality of browsing which consists of following neighbourhood links between different media, such as text and audio, rather than playing back the meeting sequentially.

Space-Based Media (or Static Media): A class of media for which space is the main structuring element. Data conveyed through space-based media are generally of a permanent and serial nature. Examples include: text and static graphics.

Temporal Neighbourhood: A set of temporal relations between space- and time-based media. In this chapter, specifically, a set of co-occurrence relations between text and speech.

Time-Based Media (or Continuous Media): A class of media for which time is the main structuring element. Data conveyed through time-based media are generally of a transient and parallel nature. Examples include: audio and video.

ENDNOTE

[1] The order of authors is alphabetical. The authors contributed equally to this chapter.

Chapter LXIII
A Proposed Tool for Mobile Collaborative Reading

Jason T. Black
Florida A&M University, USA

Lois Wright Hawkes
Florida State University, USA

ABSTRACT

This chapter presents a tool for collaborative e-learning using handheld devices that incorporates pair communication via text and speech input. It discusses the current state of e-learning for mobiles and illustrates the lack of such tools in reading comprehension domains. It then describes the tool development as a model for interface design, communication strategies, and data manipulation across mobile platforms. It is argued that such a tool can enhance e-learning among children, due to freedom of movement and variety of input (text and speech). The design is centered on a proven paper-based collaborative learning methodology which should strengthen its effectiveness. A paper prototype test that assisted in determining optimum interface layout and confirming that speech input was preferred among children is described. The system was developed and designed using creative strategies for interface layout and data manipulation. Lessons learned and plans for additional research are discussed.

INTRODUCTION

Collaboration is an important aspect of today's educational learning environment, and the infusion of technology has given rise to various studies in the area of computer-supported collaborative learning (CSCL), computer-supported collaborative work (CSCW), and computer-supported intentional learning environments (CSILE) (Jones, Dircknick-Holmfield, & Lindstrom, 2005; Scardamelia & Bereiter, 1996). The systems developed through these studies have been effectively implemented to produce major gains in comprehension of material in the math and science curriculum, but have yet to explore these benefits when applied to domains which are not math and science. The investigation of how to efficiently apply emerging technology in such environments is resulting in innovations in a wide range of systems and platforms, including handheld computers and other mobile devices.

One of the disciplines that could benefit significantly from such advancements is reading comprehension. At the present, it is apparent that reading comprehension has emerged as a major problem area in American society (Vaughn, Klingner, & Bryant, 2001). It is important to note that there are several reading comprehension tools available for the desktop platform, but the problem becomes enormous when attempting to transfer such applications to the handheld platform. There are many obstacles that must be overcome, such as limited screen real estate, smaller memory capacity, smaller processing power, and limited and often more difficult input mechanisms (such as stylus and virtual keyboard). These obstacles have led developers to steer away from this handheld platform and instead focus on the more common personal computer environment. Yet, research is indicating that the handheld computer is becoming a more viable and attractive platform due to the smaller cost, portability, durability, and increasing advancements in wireless technology (Soloway & Norris, 1999). Additionally, many scientists are investigating more innovative ways to utilize this technology and make it much more readily available to children from diverse backgrounds (MIT Media Lab, 2006).

Question-answer relationships (QAR) is a very successful learning methodology for developing reading comprehension skills (Royer & Richards, 2005; Outz, 1998; Raphael, 1986). QAR has been beneficial to educational research in that QAR not only has demonstrated the ability to improve comprehension skills of student participants, but has also shown effective implementation of peer-assisted learning strategies. There has not been a significant effort to place QAR in a computerized reading environment, and it is worth investigating whether applying QAR to a handheld learning environment would produce a more efficient reading comprehension software platform.

Thus, this chapter makes the case for collaborative reading comprehension on a mobile platform by illustrating the absence of current research in this area, describes a paper-prototype study for an interface model for collaborative reading comprehension, and then presents a handheld tool supporting collaborative reading using text and speech communication. The tool is designed using QAR as a foundation, and presents a model for development of such systems on mobile platforms. An emphasis is placed on speech input, which can further increase the robustness of user input and collaboration as a result, particularly when implemented for children.

RELATED WORK

Mobile Collaboration in Learning Environments

The explosion of mobile learning (m-learning) in educational environments is largely due to the massive influx of these portable devices in society, and more directly, in the classroom. Mobile learning takes place when users communicate wirelessly via handheld devices (phones, Personal Digital Assistants (PDAs), tablets, etc.) in the process of learning—in other words, learning that takes place with the aid of handhelds (Attewell, 2005). And, since collaboration is a natural and significant extension of a robust learning environment, it is natural to consider ways to facilitate mobile cooperation in learning activities. The mobile environment is rich with a plethora of communication tools (chat, instant messaging, shared workspaces, e-mail, and voice input/output) that make collaborative work a simple and efficient endeavor (Issacs, 2002). It is essential for researchers to explore a wide range of scenarios employing these tools in an effort to improve student learning outcomes. This research takes a look at one such endeavor.

Question-Answer Relationships

Several programs have been implemented that have shown significant development in reading comprehension skills. Among the most successful is Question-Answer Relationships (Royer

& Richards, 2005; McIntosh & Draper, 1996; Raphael, 1986), which has been shown to be a particularly effective supplement to a classroom reading program. Question-Answer Relationship teaches students to read by recognizing relationships between questions and possible sources of information, either in the text or in the reader's background. In this technique, readers are asked to read a passage and answer questions about what was read. Then, readers are required to identify the category to which each question belongs: Right There Questions (answer is explicitly in the text), Think-and-Search Questions (answer is implicitly in the text), the Author and You Questions (the answer requires you to use inference to arrive at the answer), and On Your Own Questions (the answer is entirely based on your background knowledge). Several studies have shown that students were capable of generating and answering questions that enhanced their comprehension and led to independent processing and development of knowledge (Royer & Richards, 2005; Outz, 1998). Yet, these approaches have not been incorporated in a desktop or handheld reading comprehension learning environment. It is worth investigating whether doing so will create an electronic comprehension tool which can reinforce through practice, techniques introduced by a human teacher, and hence address these issues (Vaughn, Klingner & Bryant, 2001).

Speech Recognition in Mobile Environments

This increase in the use of mobile devices has created an environment where various types of users are interacting, and as a result, researchers must utilize the full suite of modalities (or modes of input) to facilitate communication (Nanavati, Rajput, Rudnicky, & Siconni, 2006). Almost all mobile devices are equipped for voice input, making speech recognition a viable means of capturing data. In many cases, to compensate for the limited memory and power on these smaller devices, a form of *distributed speech recognition* is implemented (Schmandt, Lee, Kim, & Ackerman, 2004). In such an environment, speech is captured on the mobile device and sent across a wireless network to a server, where processing is done. The translated text is then returned to the device for use. While there are many issues to consider when utilizing this strategy (such as quality of speech, network traffic, noise, etc.), the scope of this work is to present an interface mechanism for facilitating voice input in a mobile collaborative learning session.

OVERVIEW OF SYSTEM DEVELOPMENT

Discussion of Paper Prototype Testing

Paper (also called "low-fidelity" or "lo-fi") prototyping is an interface development strategy that utilizes paper-based designs of the system and interactions with potential users with such system to arrive at an optimum design plan (Snyder, 2003). In paper prototype testing, users are asked to interact with the paper-based interface on a series of popular system tasks, with a designer playing the role of "Computer." The "Computer" mimics the actions and sounds of the system while the user progresses through these tasks, and user choices and behavior are recorded. After the session, the user is questioned in order to learn his or her cognitive processes in making decisions and the results of these answers are used to design and implement the user interface, complete with modifications indicated in the test. The attempt here is to create a "living" prototype—one that is changing to better fit the designs and recommendations of the testers involved—in order to eventually obtain the optimum design methodology for all involved. Researchers have demonstrated the benefits seen in the application of paper prototype testing—preemptive user feedback (changes are suggested before development has begun), rapid iterative development (changes can be incorporated "on the fly"), and enhanced developer/user communication (Snyder, 2003).

Participants in Paper Prototype Test

To obtain a model for interface development, a paper prototype study was conducted (Black, Hawkes, Jean-Pierre, Johnson, & Lee, 2004) involving elementary school students from a local after-school center. Five students were selected based on their background with computers (two had had experience with handhelds, one had moderate experience, and two had no experience), and age (two were in grade 2, two were in grade 3, and one in grade 4). This number of subjects is consistent with Snyder's recommendations of effective numbers of subjects in such tests, which is recommended to be between five and seven (Snyder, 2003; Nielson and Landauer, 1993). Subject #1 was a fourth grader who was a good and constant reader, and was the only subject that had familiarity with a PDA, though not much exposure. Subject #2 was a third grader who had some experience with computers, but was not a strong reader. Subject #3 was the youngest of the group, a second grader with very little computer experience, but was a strong reader. Subject #4 was a third grader who had very little computer experience and was also not a very strong reader. Subject #5 was the oldest of the group, a fourth grader who had computer experience, but was having trouble reading at grade level. All of the students had some exposure to the basic features of a computer application—buttons, passwords, pointers, and so forth—and were eager to participate in the study.

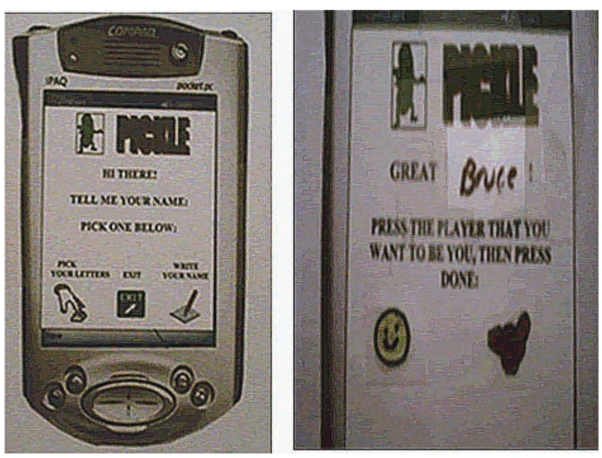
Apparatus

The test was designed using the iPAQ ™ PDA as a model (see Figure 1). A picture of the device was taken, and then scanned and printed, so that the actual size and shape of the PDA could be used in the testing. Then, cut-outs of screens to be presented (as well as buttons, menus, scroll bars, etc.) were designed and used as interchangeable interface components to be presented to subjects during completion of tasks.

Test Design

Screen mock-ups of five basic tasks were created: 1) logging into the system and selecting a partner (for collaboration), 2) reading a story, 3) answering questions, 4) e-mailing the teacher for help, and 5) chatting with their partner. Researchers participated in the test in the roles of the computer (one person transitioned screens as the computer would), and observers (who took notes on user actions and tendencies). Subjects were tested in 30 minute sessions, with two tests conducted on a given day, as recommended by Snyder (2003). During a test session, each subject was introduced

Figure 1. Image of mock-up of screen used in Paper Prototype test



to the concept of the test and why it was being conducted. They were then seated at a desk with the "computer" present as well, and the observers looking on.

Each subject was then asked to complete each of the tasks listed, with the question-answer exchanges following each task. After the completion of all five tasks, the subjects were thanked for their participation and were free to leave. Notes from observers were discussed between tests, and modifications were proposed for future tests.

Test Results

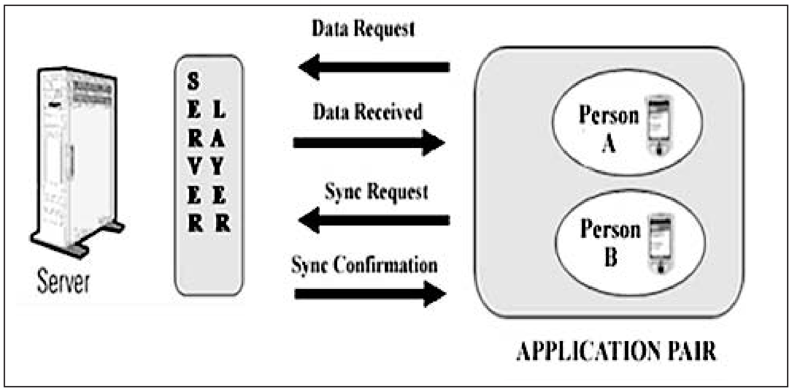
During testing, three of the five subjects preferred writing on the screen for input as opposed to the other presented forms (keypad or speech). This was not surprising, since most of these students had little computer experience, and thus would be more comfortable writing (at least initially) than using the innovative input techniques presented. Of the two subjects that did choose to pick their letters using the keypad, one subject (Subject #4) had trouble navigating the keypad and began pressing buttons on the bottom of the PDA instead of the buttons on the keypad. This action resulted in "beeps" from the "computer" indicating actions that were not allowed by the system, which further confused the subject. But what was gathered from this subject was that he was familiar with the GameBoy™ handheld computer games, which use the directional keypad on the bottom of the device for manipulation of all applications. Research has shown that students' experiences with such gaming devices can be very productive design focuses for scientists developing applications for handhelds. All subjects except Subject #1 had trouble finding the icon during the player selection phase. All subjects had no trouble finding and clicking the "Done" button.

All subjects reported no problem in reading text on the small, handheld-mock up screens, and indicated that the process was enjoyable. Subjects also had no problem transitioning between screens (done by clicking NEXT and BACK buttons presented on the interface). The chat/messaging task perhaps provided the most valuable feedback. All except one subject chose to speak their message instead of the other input features (Subject #4 chose to use the keypad in all writing tasks), indicating that this will likely be a popular feature of the application. The voice input would be very helpful to younger readers, who often do not have the ability to type or write well, and would prefer an alternative to user input. However, two of the subjects were confused when faced with the submenu that appears with the speak feature, which asks them to click on the microphone to begin speaking and to click send to transmit the message. The two subjects were unsure what buttons to press and when to press send. But once this was explained, the subjects were able to complete the task. The reliance on speech for inputting validated an earlier hypothesis which suggested that due to the age of users and simplicity of the action, speech would be chosen more often by younger users (Black, et. al, 2004). The results of this test were then used as blueprints for actual screen development.

System Overview

The interface layout was developed using an application development toolkit that allows for rapid prototyping of applications for mobile devices and provides a series of emulators that can present a simulation of the application running in or on its intended device platform. Upon completion, the finished application can easily be ported to the actual target device, and run as needed. The prototype uses a client-server approach for wireless communication (SEIR-TEC, 2002), which implies that devices are served by a central access point or base station, and communicate with the central access point through the network (see Figure 2). In this configuration, users will send computation-intensive operations to the more powerful server (such as the speech recognition) for remote processing and then download application-specific tasks to the handhelds (Omojokun, 2002). And, when usage has completed, the handhelds synchronize with the server, uploading user session information back to the server for storage in the database.

Figure 2. High-level system architecture



System Design

The process flow of the application is constructed on the question-answer relationship collaborative reading model mentioned earlier. In implementing this model, students read a passage on the handheld display and then are asked to answer questions which are downloaded from the server about the text just read. Once questions are answered correctly, students must then identify the type of questions that are presented (Right There, Think and Search, Author and You, or On Your Own), based on Raphael's Taxonomy of Questions (Raphael, 1982). Students work individually on reading and question-answering tasks, but are allowed to collaborate during periods of reaching a consensus on the correct answer (collaboration is "turned on" when questions are answered incorrectly). Students communicate by chatting with their partner as needed, as well as utilizing a shared workspace for group reflection. Students can also record personal notes in the personal journal as they progress through the lessons, as well as interact with the instructor through e-mail. All of the activities will be done on the PDA, with lessons downloaded wirelessly from the server to the PDA as requested, and student progression data being stored on the system server.

The application makes use of the standard high-level interface components—forms for user fill-in, canvases for drawing and painting of both text and images, and checkboxes and radio buttons that register user action. The screen is also touch-sensitive, allowing for stylus input at various points on the interface. Textual input is handled via stylus, keypad, and both the letter recognizer and transcriber (written text using the stylus), and is processed by device standards. The letter recognizer allows users to, after a training period, write letters using the stylus, which are "recognized" by the system and translated to their typed form. The transcriber works similarly, with users writing letters on a text pad, these letters are converted to their typed format. Both have significant learning curves on their usage, but once mastered can serve as very convenient input techniques. Figure 3 shows the initial user login screen (which is essentially a form with text fields and images that behave as buttons), a screen for selecting an icon (which uses images painted on a canvas), and a screen for registering that the user is ready to begin (again, images painted on a canvas).

These screens also make use of the standard mobile menus, which appear at the bottom of the screen just above the device soft buttons. Users can activate these menus either by clicking on them with the stylus or pressing the corresponding soft button.

The story and question screens are similarly done, with images and radio buttons dominating the device display. The user is also presented

Figure 3. Introductory screens for user input

Figure 4. Story and question-answering screens

with icons at the bottom that allow him or her to activate the various system functions (using the *Diary* for personal reflection; sending a message to the instructor, using the *Group Workbook*, etc.). These are illustrated in Figure 4.

Collaboration Components

The collaborative features implemented in the prototype system are shared workspaces, e-mail messaging, personal reflection, and chat services. The shared workspace is implemented as a *group workbook* that is visible to both participants of a team. Each user sees an up-to-the-second image of the workbook and changes made (by entering data in the workbook) are broadcast to each user's device. The system synchronizes access to this feature, locking it while it is being written to so that changes can be implemented before additional writes are allowed. The workbook enables each user to jot down notes that may assist the team in answering questions in later sessions. These notes can be entered either by keypad or writing (textual) or by speaking the text (voice) (see Figure 5).

The personal journal is used for reflection as the user moves through sessions. Each journal is seen only by that user and is updated upon request. All additional entries into the journal are added to its previous contents, similar to writing in a paper journal. The entries are recorded by being sent to the server for storage upon completion of the session period. Upon the next login, the cur-

Figure 5. Workbook and diary screens

rent contents of the journal are sent to the client should any new entries be desired. As with all other methods of input, the user can provide the journal entry either through textual or voice input (see Figure 5).

The system allows users to send e-mail to the instructor in the event that assistance is needed. The sending of the message is implemented by the client sending a message to the server and the server forwarding that message to the instructor's e-mail address. A record of the transmittal is also stored by the server for reporting purposes. Again, the message can be either in the form of text or voice input (see Figure 6).

Chat Service

The chat service is implemented similar to the standard chat service hosted by any Internet service provider. The system registers that a user wishes to chat and sends a message to the user's partner that he or she wishes to chat. Once a confirmation has been received by the partner, the chat session begins, with users typing in messages (or entering them via voice input) and those messages being displayed on the screen. These messages are also recorded by the server for reporting. When chat is no longer desired, the user indicates this and the session resumes from its previous point. The chat provides the users the opportunity to reach a consensus on certain learning tasks and facilities the peer-tutoring methodology, both techniques present in successful collaborative environments (see Figure 6).

Speech Recognition Strategy

Since the system is designed to be adaptable to a variety of environments, it may be the case that the keypad, recognizer, and transcriber are too complex in a setting of younger users. Thus, the system also allows for Speech Input, where users are allowed to speak their messages into the system, and these messages are converted to text and displayed on the screen (or sent via e-mail if needed). This is done to take into account that younger users may not be good typists or even know how to spell well, but may still wish to enter data.

The ability to provide speech input for users is a major component of the architecture. The current literature does not indicate any use of this feature in studies involving handhelds and collaboration. In the math and science-centered applications, input is often simple, with users asked to select items or to enter numbers as part of equations. This poses a problem in non-science domains, where input may often be sentence-structured and much more verbose. There needs to be an additional method of providing this type of input, and speech or voice input fills this need adequately.

Figure 6. Chat and help screens

The application environment implements a strategy for dealing with speech input and/or speech output. The system receives the data sent to the server and runs it through a speech recognizer program to produce written text. The written text is then sent back to the application to be displayed on the screen. The server is responsible for handling speech requests sent by the Diary, Help, Journal, Workbook, and Chat applications and funneling them to the appropriate mechanism.

DISCUSSION

It is important to note that while this work demonstrates that this type of interface can indeed be developed for mobile devices, the actual testing of this system in a live classroom will occur in additional studies. The researchers are currently working with school teachers to develop a curriculum model that can incorporate such a system, in an effort to determine if its application would be effective in improving reading skills of younger students. This is a daunting task, but one that holds much promise for both computer scientists and educators as well.

Prior to implementation of a complete system, there are issues to be examined related to data management, data modeling, logging of system and user operations and functions, and server-side management. There are also issues regarding data security and reliability of data to consider. And, while the screen size is likely appropriate for beginning readers and younger users, there may be an issue regarding displaying of material on the device for more advanced users. The current model only displays one page of data at a time, and pages are turned and not scrolled. More advanced users would likely want to remain on a page and simply scroll down or up to view additional material. This would call for some device other than the PDA, or at the very least, in a revision of the type of screen layout. However, the focus of this project is younger children, and thus the designed system is very appropriate.

CONCLUSION AND FUTURE WORK

The development of this collaborative mobile learning system and its implementation on the actual target handheld devices is an indication that this type of architecture is possible and proved both challenging and rewarding. Each of the desired screen designs was capable of being constructed in the chosen language and the interaction between screens was simple to maintain. The communication between devices and between device and server was easily maintained via a wireless network and access to a server machine. Using a Web server allows for testing and demonstration of the system in any environment where wireless Web access is available.

The implementation of QAR in a collaborative platform was also successful. As mentioned earlier, QAR requires individuals to work in pairs, which is accommodated by the interface in this system. QAR also expects students to not only read, but also answer questions and then identify categories of questions. The multiple screens developed in this system also accomplish this task. And, since QAR is an extremely successful tool (in a paper-based environment) in enhancing reading comprehension skills, it is rewarding to note that the interface presented does not take away from the functionality and robustness of the methodology, but stresses it very well.

Speech input is a very significant feature of any collaborative environment and the tools included in the system provide for that capability. Students are able to speak words of communication with partners, and these words are indeed translated and presented on the screen. This is a major component since mobile devices often have challenging input techniques (using a stylus for large volumes of text can be very cumbersome). This system addresses and solves this issue as well.

The next step is to implement this system in the actual classroom, with the assistance of grade school teachers and administrators, in the effort to study its effect on reading comprehension skills. It is believed that students using this system will become better readers, and that the system's integration into the classroom learning setting will be unobtrusive and seamless. Since most reading comprehension software is developed for the desktop environment, utilization of such a system in this mobile platform could prove very exciting and rewarding, serving to fill a much needed void in the collaborative learning spectrum.

Overall, the system presented in this work provides one possible approach to developing collaborative learning environments on intermittent devices, successfully providing an architecture for modeling interfaces for smaller, more limited machines. This research is just scratching the surface of what is capable for reading comprehension software, showing that tomorrow is promising for addressing the crisis of improving children's reading skills nationally.

REFERENCES

Attewell, J. (2005). Mobile learning: Reaching hard-to-reach learners and bridging the digital device. In G. Chiazzese, M. Allegra, A. Chifari, & S. Ottaviano (Eds.), *Methods and technologies for leaning.* (pp. 361-365). Southampton: WIT Press.

Black, J., Hawkes, L., Jean-Pierre, K., Johnson, I. & Lee, M. (2004, September 13-16). A paper prototype study of the interface for a children's collaborative handheld learning application. In *Proceedings of Mobile HCI 2004*, Glasgow, Scotland.

Isaacs, E., Walendowski, A., & Ranganathan, D. (2002). Mobile instant messaging through Hubbub. *Communications of the ACM, 45* (9), 68-72.

Jones, C., Dirckinck-Holmfeld, L., & Lindström, B. (2005). CSCL The next ten years—A view from Europe. In T. Koschmann, D. Suthers, & T-W. Chan (Eds.), *Computer supported collaborative learning 2005: The next ten years!* Mahwah, NJ: Lawrence Erlbaum Associates.

McIntosh, M.E., & Draper, R. J. (1996). Using the question-answer relationship strategy to improve students reading of mathematics texts. *The Clearing House*, *154*, 161.

MIT (2006). *MIT Media Laboratory*. Retrieved from http:// www.media.mit.edu/about/overview.pdf.

Nanavati, A., Rajput, N., Rudnicky, A. I., & Sicconi, R. (2006, September). Workshops: SiMPE: speech in mobile and pervasive environments. In *Proceedings of Mobile HCI 2006*. Helsinki, Finland.

Nielsen, J., & Landauer, T. K. (1993). A mathematical model of the finding of usability problems. In *Proceedings of Conference on Human Factors in Computing Systems INTERCHI '93*, Amsterdam (pp. 206-213). New York: ACM Press.

Ouzts, D. (1998). Enhancing the connection between literature and the social studies using the question-answer relationship. *Social Studies & The Young Learner, 10*(4), 26-28.

Raphael, T. (1982). Question-answering strategies for children. *The Reading Teacher, 36,* 186- 19.

Raphael, T. (1986). Teaching question answer relationships, revisited. *The Reading Teacher, 39,* 516-522.

Royer, R., & Richards, P. (2005). Revisiting the treasure hunt format to improve reading comprehension. Retrieved from http://www.iste.org/Content/NavigationMenu/Research/NECC_Research_Paper_Archives/NECC_2005/Royer-Regina-NECC05.pdf

Scardamalia, M., & Bereiter, C. (1996). Student communities for the advancement of knowledge. *Communications of the ACM, 39*(4), 36-37.

Schmandt, C., Lee, K. H., Kim, J., & Ackerman, M. (2004, June). Impromptu: Managing networked audio applications for mobile users. In *Proceedings of MobiSys 2004* Boston, Massachusetts.

Snyder, C. (2003). *Paper prototyping: The fast and easy way to design and refine user interfaces.* San Francisco: Morgan Kaufmann Publishers.

Soloway, E., & Norris, C. A. (1999). *Schools don't want technology, schools want curriculum*. Retrieved in 2003 from http://www.cisp.org/imp/june_99/06_99soloway-insight.htm

SouthEast Initiatives Regional Technology in Education Consortium (SEIR-TEC) News-Wire. (2002). *Using handheld technologies in schools, 5*(2).

Trifonova, A., & Ronchetti, M. (2005). Prepare for a bilingualism exam with a PDA in your hands. In G. Chiazzese, M. Allegra, A. Chifari, & S. Ottaviano (Eds.), *Methods and technologies for learning* (pp. 343-347). Southampton: WIT Press.

Vaughn, S., Klingner, J. K., & Bryant, D. P. (2001). Collaborative strategic reading as a means to enhance peer-mediated instruction for reading comprehension and content area learning, *Remedial and Special Education, 22*(2), 66-74.

KEY TERMS

Collaborative Learning: An environment where students work alone or in groups to complete a set of tasks, usually lessons, where they assist each other in learning.

Computer-Supported Collaborative Learning (CSCL): The study of users collaborating in a computerized environment on learning tasks.

Computer-Supported Intentional Learning Environments (CSILE): Database software that provides tools for organizing and storing knowledge as a means of sharing information and thoughts with peers, supporting both individual and collaborative learning.

Computer-Supported Collaborative Work (CSCW): The study of how people work with computers and how they can work with each other using them.

Distributed Speech Recognition: The process of capturing speech on a mobile device and transporting it via wireless network to a server to be processed ("recognized") or translated, and subsequently returning the translated speech to the mobile device.

Mobile Learning (M-Learning): Users communicating wirelessly via handheld devices in the process of learning.

Question-Answer Relationships (QAR): An instructional methodology for enhancing reading comprehension skills by teaching students to answer questions and generate their own questions based on text.

Chapter LXIV
Evaluating Learner Satisfaction in a Multiplatform E-Learning System

Tiong T. Goh
Victoria University of Wellington, New Zealand

Kinshuk
Athabasca University, Canada

Nian-Shing Chen
National Sun Yat-Sen University, Taiwan

ABSTRACT

The main objective of this chapter is to present a comparative evaluation between two e-learning systems from the end user (learner) perspective. The evaluation instrument is based on a multiplatform e-learning systems framework and a modified version of the Questionnaire for User Interface Satisfaction (QUIS). First, the evaluation intends to compare the achievable level of overall the learner satisfaction score between a Blackboard e-learning system and a multiplatform e-learning system with three different accessing devices. Second, the evaluation intends to explore the degree of influence and identifies grouping relationships among the factors that influence learner satisfaction while engaged in a multiplatform e-learning system. Lastly, the evaluation determines the gain in the learner satisfaction score between the two systems with respect to three different accessing devices. The findings and the process of evaluation can play an important role for the designer to improve the adaptation process and to enhance the level of learner satisfaction in future multiplatform e-learning systems.

INTRODUCTION

With the advent of mobile devices and wireless infrastructure, mobile devices can provide learning opportunities anytime and anywhere. However, using mobile technology for e-learning is not widely adopted. One factor is the adaptation of content from PC to mobile devices (W3C, 2001). In e-learning systems, content adaptation not only requires the layout, format, and structure of the content to be modified, it is also compounded by the problem of adapting interactive multimedia content. Interactive multimedia content is less likely to be accessible via mobile devices without appropriate content adaptation. Even if the content has been adapted for different devices, it is still not clear how adaptation and other factors influence learner satisfaction. However, without appropriate adaptation for various accessing devices, adoption of technology is less likely according to the technology acceptance model (Davis, 1989, 1993).

This chapter first provides a brief review in content adaptation techniques and W3C independence activities (DIA). Subsequently, the chapter discusses the multiplatform e-learning systems framework. It should be stressed that the formulation of the multiplatform e-learning systems framework involves other important research areas such as device identification, bandwidth estimation, e-learning system architecture and frameworks, and many more. Due to the focus of the chapter, only work on content adaptation and W3C independence principles were reviewed. Subsequently, a comparative evaluation is presented on a multiplatform e-learning system and a Blackboard e-learning system to assess learner satisfaction with three accessing devices. The Blackboard e-learning system is a popular learning management system adopted by many universities (Blackboard, 2006). The results and findings based on statistical analysis are presented. These findings and the process of evaluation can play an important role for the designer to improve the adaptation process and to enhance the level of learner satisfaction while engaging in multiplatform e-learning systems. Finally, limitations and recommendations for future research conclude the chapter.

BACKGROUND

Content Adaptation

One of the main issues for a multiplatform e-learning system is the requirement to provide meaningful access to an increasing number of accessing devices. Most Web-based systems including e-learning systems use server-side designs that are oriented towards clients with standard screens and standard HTML browsers such as a PC. These designs are unlikely to be acceptable for personal mobile devices such as mobile phones or PDAs. In order to provide meaningful access and learning, the content (including multimedia), structure, and navigation must be able to adapt. Thus a multiplatform e-learning system needs adaptivity, that is, the ability of a system to adapt itself to different client needs. For the case of a multiplatform e-learning system, this includes the ability to cope with bandwidth, memory and power limitations, restricted presentation capabilities, and different user profiles. Content adaptation techniques are examined first to understand and appreciate how different techniques can be helpful in the multiplatform e-learning system implementation process. One should aware that some of these techniques, because of their inherent assumptions, may not always be suitable for the current implementation.

The Adaptive Web Content Delivery (AWCD) framework (Chen, Yang, & Zhang, 2000; Ma, Bedner, Chang, Kuchinsky, & Zhang, 2000) provides a useful overview of the problems that have to be addressed in adaptive systems. Firstly, the system must be able to detect the parameters needed for adaptivity. This includes the automatic measurement of the network bandwidth and the loading and registration of used end-devices and user preferences. One approach to the latter is to use Web forms. More sophisticated inference mechanisms based on user behaviour patterns can also be adopted. And secondly, the system must provide a decision engine for determining when and how to adapt content. In the multiplatform e-learning system, an XML document string for each factor was adopted. The advantage of using an XML string is the ability to provide simple exchange mechanisms with other modules and the potential to transform into other developing standards if necessary. The second issue is included in the coordination dimension.

In general, there are two different classes of approaches to content adaptation. The first class of approaches deals with the problem on the level of pages, that is, that the Web-based service is fully implemented by Web pages that can be displayed by a standard browser on a standard computer display, is assumed. Then adaptation has to take HTML as the format of its input. In the UWA framework (Finkelstein et al., 2002; Kappel et al., 2002; UWA, 2002), the adaptation can be expressed by event-condition-action rules. This approach assumes a design has already been made in the PC environment and the adaptation is to transform the PC Web page into another device environment. This assumption is not necessary in the multiplatform e-learning system development. There is no necessity to assume PC Web pages as a starting point but one should recognise that many existing e-learning systems do. Therefore, some of the disadvantages of the page level adaptation approach may not be applicable in the multiplatform e-learning system as envisaged.

The second class of approaches handles the problem on a conceptual level, that is, assuming there will be a generic abstract description of the service's content, from which the actual Web pages could be generated. Adaptation now takes the abstract description as its input and derives another abstract description. The pages generated out of this modified specification can be more suitable for specific clients, in particular mobile clients. This approach may not allow other factors such as preferences and interactivity to be optimally incorporated into the cost function (Tretiakov & Kinshuk, 2004) which may affect the multiplatform e-learning system implementation.

For content adaptation on the level of pages, the most prominent approaches are re-authoring (Bickmore & Shilit, 1997) and transcoding (Bharadvaj, Joshi, & Auephanwiriyakul, 1998; Hori, Kondoh, Ono, Hirose, & Singhal, 2000; Smith, Mohan, & Li, 1999). Re-authoring applies functions to Web page descriptions, which will result in new descriptions that are better suited for various kinds of clients. Transcoding aims at direct manipulation of the HTTP stream. Based on these techniques, several commercial products and system prototypes have been implemented such as Digestor (Bickmore & Shilit, 1997), Spyglass (Spyglass, 1999), Intel QuickWeb (Intel, 1998), Mobiware (Angin, Campbell, Kounavis, & Liao, 1998), TranSend (Fox, Gribble, Chawathe, & Brewer, 1998), WingMan (Fox, Goldberg, et al., 1998), and Power Browser (Buyukkokten, Garcia-Molina, Paepcke, & Winograd, 2000).

Adaptivity on the level of page descriptions can be achieved more easily if data content, navigation links, and presentation are separated from each other, for example, by using XML, XLink, and XSL (Goldfarb & Prescod, 1998). A multiplatform e-learning system experience suggests that content, navigation, presentation, and logic should be separated from each other. Also, the work of Chen, Zhou, Shi, Zhang, and Wu (2001) presents an interesting approach to detecting objects in pages and their categories, using a functional object model. Adaptation rules are then applied to these objects. The rationale is that every object in a Web site serves certain functions, which are either basic or specific functions. FOM distinguishes between objects, which themselves can be basic or composite, and object categories. Based on this, a complete Web site is transformed first into a FOM model, before adaptation rules are applied to the model rather than to the Web pages.

Besides page-based techniques, newer techniques concentrate on content adaptivity at the conceptual level (Feyer, Kao, Schewe, & Thalheim, 2000). One such technique is the media type approach. Media types were introduced by Feyer et al. (2000) as a means to formalise Web-based systems. The major intention was to abstract completely from presentational issues, but nevertheless, to keep in mind the capability for presentation. Media types provide a conceptual abstraction from the page level based on extended views. One of these extensions addresses adaptivity. The underlying idea is to include information that indicates which parts of media type content should preferably be kept together and which parts can be separated. Two mechanisms have been developed for this, one based on sub-typing, and the other on numeric proximity values. The apparent difference between the two classes of approaches is that for page-level adaptivity, the adaptation rules will become part of the Web-based service, whereas adaptivity on a conceptual level must provide generic rules for adaptivity. Application-dependent adaptivity may be easier to achieve, but it is expensive with respect to development and maintenance.

W3C Device Independence Activities

W3C device independence activities (DIA) (W3C, 2001) is closely related to the research in multiplatform e-learning system frameworks. The goal of the DIA is to develop ways for future Web content and applications to be authored, generated, or adapted for a better user experience when delivered via many device types. In DIA, user experience is defined as a set of material rendered by a user agent, which may be perceived by a user and with which interaction may be possible. The goal of the multiplatform e-learning system framework is to devise a framework for accessing e-learning content and for learning to take place via different accessing devices with the aim of analysing the level of influence of the factors within the framework. The current research therefore complements DIA by focusing on basic questions relevant to e-learning and learner satisfaction. In addition, the current multiplatform e-learning framework takes insights from the DIA perspectives of user, authoring, and delivery mechanisms to formulate a competency framework that focuses on content, learner, device, communication, and coordination dimensions that better fits the e-learning environment.

Motivated by future access scenarios, DIA looks at how Web content can be accessed from three perspectives: the user perspective, the author perspective, and the perspective of delivery mechanism. DIA identifies seven working principles to achieve the goal (W3C, 2001). The seven working principles are: device independent access, device independent Web page identifiers, functionality, incompatible access mechanism, harmonisation, characterisation of delivery context, and adaptation preferences. The goal of the device independent access and device independent Web page identifier principles is to ensure that a functional user experience is always possible via any access mechanism. In the multiplatform e-learning system implementation, this is achieved by matching the devices' capabilities with functional content. In the evaluation process, comparison is made between e-learning systems that provide functional user experience to various accessing devices to determine the level in learner satisfaction. The functionality, incompatible access mechanism, and harmonisation principles aim to ensure that functional experience, if not met, should give appropriate feedback to the user and that harmonised user experience is possible from the author perspective. A harmonised user experience is one that meets the user delivery context and also the quality criteria of the author. In the multiplatform e-learning system, if the user delivery context can not be met, the lowest version of the functional user experience is rendered, depending on the device's capability. The harmonised experience is governed by the coordination dimension which provides harmonised adaptation based on attributes from other dimensions such as bandwidth, device capabilities, and environment factors. Finally, the characterisation of delivery context and adaptation preference principles ensures that delivery context is made available to the adaptation process and that users can change their preferences to modify the adaptation process. In the multiplatform e-learning system, the delivery context such as bandwidth, device profiles, learner profiles, and preference are either stored in XML files or estimated in real time to allow the adaptation to process. User preferences can be changed by the user to be reflected in the adaptation process.

The device independence activities mentioned strongly demonstrated a vision for the future of a multiple devices accessing scenario in general Web service. The current research of a multiplatform e-learning system framework is aligned with the direction of W3C's DIA. This research concentrates on the domain of e-learning and also devises a set of competency principles to operationalise the multiplatform e-learning system framework. E-learning is a revolutionary way of teaching. With the advent of mobile technology, e-learning is constantly evolving, thus there is still a need for more research. The multiplatform e-learning research also aims to gain new knowledge by focusing on fundamental factors influencing learners when exposed to a variety of accessing devices. In this regard, the current research goes beyond and complements the DIA activities.

MULTIPLATFORM E-LEARNING SYSTEMS

A multiplatform e-learning system is a system that can be accessed by different types of devices such as a PC, a PDA, and a mobile phone. A multiplatform e-learning system has the adaptive capability to deliver appropriate content depending on the accessing devices type, learner profiles, and the context of use. In order to address the issues related to a multiplatform e-learning system, a framework for a multiplatform e-learning system is constructed.

Multiplatform E-learning Systems Framework

The multiplatform e-learning systems framework intends to serve three purposes. First, it provides a starting point for requirement analysis similar to the information system architecture (Zachman, 1987). It serves as a template that allows system analysis to capture the requirement. Second, it serves as a guideline for implementation. A framework is useful only if it can be operationalised. The adaptation framework for a multiplatform e-learning system has been operationalised to validate its usefulness via an implementation. Third, the framework serves as a test bed for comparative evaluation of the Blackboard e-learning system and the multiplatform e-learning system. The adaptation framework (Goh & Kinshuk, 2004) also describes the competency components within an e-learning system where the learner could engage using multiple platforms. Five core dimensions have been identified: content dimension, learner dimension, device capability dimension, communication dimension, and coordination dimension. While some of these dimensions may look genetically similar to traditional e-learning systems, the adaptation perspective, multiple devices requirements, and context of use such as learner dimension and device dimension would differ in operation. Each of these dimensions includes various sub-dimensions as depicted in Figure 1. This section intends to explain some of these sub-dimensions that constitute to the adaptation framework. In addition to identifying the core dimensions, the framework also describes the required competency needed for adaptation with respect to multiplatform engagement scenarios. It should be stressed that the formulation of the adaptive framework for a multiplatform e-learning system is derived from extensive literature review. The literature review covered multidisciplinary areas related to e-learning systems framework and architecture, content adaptation, device identification, mobile learning system, context awareness system, RDF, CC/PP/UAProf, MIDF-Mobile information device profile, W3C device independence access, IS theory in information system architecture, agent-oriented systems, ITS systems, Web service coordination, and coordination theory.

Content Dimension

Content dimension represents the actual context and knowledge base of the e-learning system. It includes various sub-dimensions. The organisation sub-dimension describes how content presentation be organised. Attributes such as modular, hierarch, and mixed structure are the main consideration. Another sub-dimension is the granularity level of the content that indicates the level of difficulty of the content presented to the learner. Learning module sub-dimension represents the learning object of the actual material normally represented by text or multimedia of the content. This includes the use of text, audio, animation, video, 3-D video, and so on, to represent the content to the learner. Pedagogy sub dimension represents the teaching models (strategies) that the system adopts.

Competency Principle 1

The multiplatform e-learning system must have the competency of selecting the appropriate organisation structure, learning content, granule and teaching strategy and deliver according to the engaging situation.

Learner Dimension

The learning profile sub-dimension of the learner dimension includes the static learning history and mobile learning history attributes. These are normally represented by parameters such as module completed dates, weight and score, time taken, date of last access, and devices access type, and so on, depending on the algorithms used in determining the learner profile. Learner preference sub-dimension contains presentation preference and learning preference attributes. These are normally represented by parameters such as preferred difficulty level, background, font, and learning style. Environmental sub dimension represents the actual social factors surrounding the learner when the system is in use. The environment can be noisy. The social environment might require self exploration without external help or even multi-tasking. Different environments such as café, hot spot, and classroom situation will have to be adopted differently. Motivation sub dimension reflects the emotional aspect of the learner, such as urgency. The adaptation must take into account of the degree of need in using the system.

Competency Principle 2

The multiplatform e-learning system must have the competency of organising, extracting and utilising the learner information to best suit the learner.

Device Dimension

Device dimension consists of the presentation capabilities sub-dimension which relates to the presentation attributes such as the media player, screen, and browser. The operational capabilities sub dimension relates to the hardware aspect of the device such as CPU, memory, and power. The ergonomic capabilities sub dimension relates to the ease of use of the device and the browser learner interface.

Competency Principle 3

The multiplatform e-learning system must have the competency of identifying and utilising some or all of these capabilities of the devices.

Communication Dimension

The communication dimension relates to the quality of the network and the delivery mode. Under this dimension, there are three sub-dimensions. The connectivity sub dimension consists of attributes such as wired and wireless connection of the learning situation. The quality sub dimension relates to the perceived response and perceived stability of the system from the learner perceptive. The delivery sub dimension relates to the mode of deliver such as real-time online mode, pre-fetching, and off-line synchronization mode.

Competency Principle 4

The multiplatform e-learning system must have the competency of deciding which mode of operation is best suited for the communication conditions.

Coordination Dimension

The coordination dimension represents the software and algorithm sub-dimension of the application, the presentation sub-dimension, the interactivity sub-dimension of the application, and the navigation sub-dimension. In any adaptive systems, these dimensions must be well coordinated to provide the learner with good learning satisfaction. The software and algorithm sub-dimension contains the script language and server page language to control the flow of the application from feedback through interactivity and navigation sub-dimensions. The presentation sub-dimension links to the display and transforma-

Figure 1. A framework for multiplatform e-learning systems

Content Dimension	**Learning module**	**Organisation**	**Granularity**	**Pedagogy**
	text	modular	beginner	teaching strategy
	multimedia	hierarchy	intermediate	(guideline)
		mixed	advanced	
			expert	
Competency Principle 1:	**The multiplatform e-learning system must have the competency of selecting the appropriate organisation structure, learning content, granule, and teaching strategy and deliver according to the engaging situation.**			
Learner Dimension	**Environment**	**Motivation**	**Learning Profile**	**Preference**
	quietness	urgency	static learning history	presentation preference
	available of help self exploration	(trendy technology)	mobile learning history	learning preference
	multitasking			
Competency Principle 2:	**The multiplatform e-learning system must have the competency of organising, extracting, and utilising the learner information to best suit the learner.**			
Device Dimension	**Presentation Capabilities**	**Operational Capabilities**	**Ergonomic Capabilities**	
	media player	cpu	platform (input)	
	screen	memory	browser	
	browser	power	portability	
		sensor	mobility	
		networking		
Competency Principle 3:	**The multiplatform e-learning system must have the competency of identifying and utilising some or all of these capabilities of the devices.**			
Communication Dimension	**Delivery**	**Quality**	**Connectivity**	
	real-time	perceived response	wired	
	synchronised (off-line)	perceived stability	wireless	
	pre-fetch			

continued on following page

Figure 1. contiued

Competency Principle 4:	The multiplatform e-learning system must have the competency of deciding which mode of operation is best suited for the communication conditions.			
Coordination Dimension	**Navigation**	**Interactivity**	**Presentation**	**Software and Algorithm**
	link	forms	display	script
	button	parameters	transform	servlet/server page
	hot spot	variables		cgi
	option	feed back		agent
	list	cookies		Web service
Competency Principle 5:	**The multiplatform e-learning system must have the competency of effectively isolating the content, presentation, navigation, and interaction components and subsequently integrating them seamlessly and effectively.**			

tion of the content to the learner. The interactivity sub-dimension represents how the user information can be sent back as feedback to the application. And, the navigation sub-dimension provides both feedback and movement within the application. For instance, in the communication dimension mentioned earlier, when the learner operates under a synchronisation sub-dimension, certain interactivities in the coordination dimension have to be dropped and cookies and script must be activated to store interactivity information such as the answers to a test. The coordination dimension here can provide these adaptations.

Competency Principle 5

The multiplatform e-learning system must have the competency of effectively isolating the content, presentation, navigation, and interaction components and subsequently integrating them seamlessly and effectively.

EVALUATION METHODOLOGY

Evaluation Objectives

As mentioned earlier, this chapter intends to present a comparative evaluation based on the multiplatform e-learning system. The evaluation intends to achieve the following objectives:

- **EO1**. Determine the achievable level of the overall learner satisfaction score between the Blackboard e-learning system and the multiplatform e-learning system.
- **EO2**. Determine the gain in the learner satisfaction score for different accessing devices in the multiplatform e-learning system.
- **EO3**. Determine the degree of influence in the learner satisfaction score among the factors in a multiplatform e-learning system.
- **EO4**. Determine any groupings among the factors that influence the learner satisfaction score in a multiplatform e-learning system.

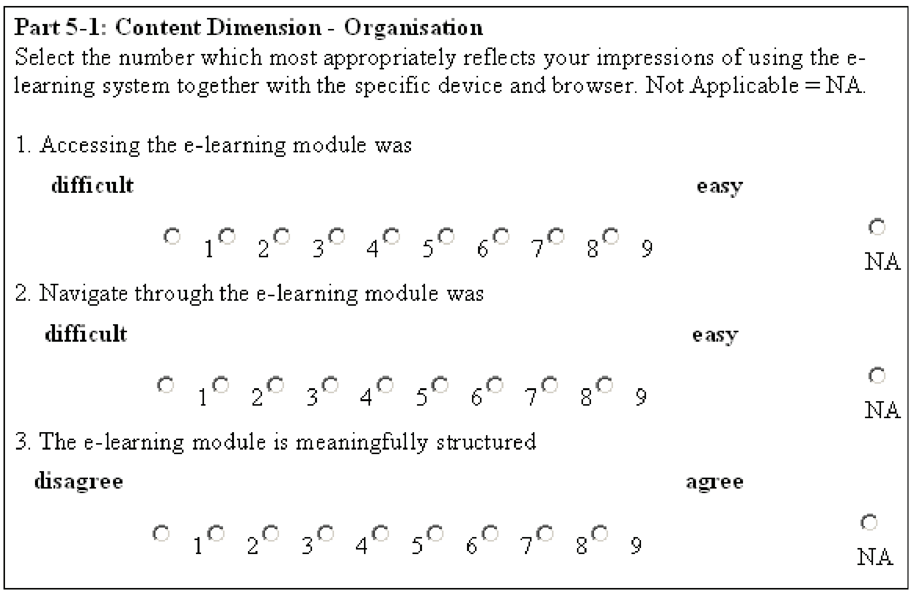
Instrument

A QUIS was developed by Chin, Diehl, and Norman (1988) at the University of Maryland. QUIS has been used for evaluating Web-based e-learning systems and comparative studies of PDA based and paper based quizzes (Johnson, Zhang, Tang, Johnson, & Turley, 2004; Segall, Doolen, & Porter, 2005). The original QUIS consisted of five scales which show overall reactions to the software, screen, terminology and system information, learning and system capabilities. The first scale relating to overall reactions to the software consisted of six questions and the other four scales include a total of 21 questions. The scaling of items ranges from 1 to 9 and an additional "no answer" option is available. Compared to other instruments, such as the Purdue Usability Testing Questionnaire (PUTQ) (Lin, Choong, & Salvendy, 1997), the Post Study System Usability Questionnaire (PSSUQ), and the Software Usability Measurement Inventory (SUMI) (Lewis, 2002), the QUIS is considered closer to the current evaluation objective. The learner questionnaire for evaluation of the multiplatform e-learning system (MELQ) in this study is modified and extended from the QUIS. This questionnaire consists of nine parts. Part 1 of the MELQ consists of capturing demographic information. Part 2 measures e-learning satisfaction. Part 3 measures previous interaction experience and was modified to include more up-to-date and relevant browsers, mobile games, SMS, PDA, and mobile devices. Part 4 measures the overall learner satisfaction score (LSS) and the statements were modified to include the e-learning context for different accessing devices. Statement number 5 was changed to reflect the meaningfulness of the system. Part 5 measures the content dimension (D1) with sub parts measuring organisation, granularity, multimedia objects, and pedagogy. Part 6 measures the learner dimension (D2) with sub parts measuring environment, available help, motivation, and preferences. Part 7 measures the device dimension (D3) with sub parts measuring presentation capability and ergonomic capability. Part 8 measures the communication dimension (D4) with sub parts measuring perceived response, perceived stability, and mode of delivery. Part 9 measures the coordination dimension (D5) with sub parts measuring navigation, interaction, and presentation. A sample of MELQ is shown in Figure 2.

Evaluations Scenarios

Evaluating a system is a difficult task and it is even more difficult when the system is adaptive (Hook, 1997). Evaluation of adaptive systems remains a challenging issue (Weibelzahl, 2002). Most studies of adaptive systems are comparisons of the systems with and without adaptivity (Hook, 1998; Boyle, Craig, & Encarnacionj, 1994; Brusilovsky & Pesin, 1994; Meyer, 1994). The

Figure 2. Sample of MELQ


Part 5-1: Content Dimension - Organisation

Select the number which most appropriately reflects your impressions of using the e-learning system together with the specific device and browser. Not Applicable = NA.

1. Accessing the e-learning module was

difficult … **easy**

1 2 3 4 5 6 7 8 9 NA

2. Navigate through the e-learning module was

difficult … **easy**

1 2 3 4 5 6 7 8 9 NA

3. The e-learning module is meaningfully structured

disagree … **agree**

1 2 3 4 5 6 7 8 9 NA

current evaluation is also a comparative study. Six scenarios were defined with the intention to vary the interaction between learners and the e-learning systems. Six participants were enlisted to take part in each scenario. A total of 36 participants were enlisted. For each scenario, each participant has to complete several specific tasks for each system with three different accessing devices. Thus, the experiment can be viewed as repeated measures with 6x2x3 (6 scenarios x 2 e-learning systems x 3 devices) unique experiments. The tasks consist of login to the respective system, navigate to the respective folder, navigate to the contents pages, explore individual learning module, explore interactive multimedia, login to exercise, response to exercise, revisit content if necessary, view exercise feedback, and exit system. E-learning system A consists of learning and assessment modules implemented in the commercial Blackboard system without adaptive capabilities. E-learning system B consists of a multiplatform e-learning system with adaptive capabilities. Participants accessed different modules on different devices. Upon completing each system with a device, the participants were requested to fill out a survey form (MELQ). Once all the tasks were completed, each participant was interviewed. A total of 216 data sets were generated from the experiments.

For scenario one, participants assumed that they were in a classroom setting and that they had just purchased an IPAQ PDA and a Nokia mobile phone. They were keen to explore these devices to learn about a particular physics topic. The scenario encouraged the participant to ask as many questions as possible during the evaluation process. The evaluation environment was made as relaxing as possible. The goal for the scenario was to achieve the best measurement score possible for each device and each e-learning system.

Scenario two also assumed a classroom setting. Participants used PCs, PDAs, and mobile phones to explore the two e-learning systems. In scenario two, participants were made to experience a low bandwidth connection. The multiplatform e-learning system detected and estimated the delay and offered the participants off-line content delivery while the Blackboard system delivered content as usual, with delay. It should be stressed that both systems experienced the same delay.

In scenario three, participants were made to experience an interruptive environment. Participants were required to read and close a randomly appearing popup window while performing the sequence of required tasks. The objective was to simulate a multi-tasking situation frequently encountered in a mobile environment. The multiplatform e-learning system made adaptive adjustments to the delivery of content to the learner by adjusting the length of assessment content while the Blackboard e-learning system remained the same.

In scenario four, participants were made to experience a busy bandwidth connection, but not as long a delay as the off-line scenario. The multiplatform e-learning system detected and estimated the bandwidth and presented a pre-fetch delivery solution while the Blackboard system delivered content as usual, but with some delay.

In scenario five, participants were assumed to be receiving a call from a friend who wanted help on a particular physics topic, and that the participant did not have much time because of an imminent appointment with a doctor. This scenario simulated the condition of working under time constraint. An 8 minute countdown clock was deployed to increase the urgency. The adaptive e-learning system provided adaptive content delivery by adjusting the length of the assessment.

In scenario six, participants were made to self explore the e-learning systems without any help from the researcher. In addition, they were warned that 10 points would be deducted for every question asked. The objective was to examine the factors influencing the learner when absolutely no help was given during self exploration. This scenario is frequently encountered when learners use mobile devices to access e-learning systems outside of a classroom environment.

It should be emphasised that in all these scenarios the multiplatform e-learning system was capable of identifying the accessing device, determining its capabilities, estimating its bandwidth, and providing content adaptation through style sheet transformation. In reporting the results in the next section, the abbreviations in Table 1 will be adopted.

PARTICIPANTS INFORMATION

A total of 36 students took part in the evaluation. Six scenarios were used to vary the interaction between the e-learning systems, accessing devices, and learners. Out of the 36 participants, 20 (55.6%) were female and 16 (44.4%) were male. The participants were mostly (80%) aged between 16 and 25, while 8% were aged between 26 and 30, and 12% between 31 and 45. The majority (89%) were doing an undergraduate degree while 11% were at post graduate level. The samples used were convenient samples and matched the university's student population's distribution and profiles.

The Internet Explorer (IE) browser usage among the students was very high: 44% of the participants spent over 10 hours a week using IE, 28% of the participants spent 4-10 hours a week, 14% of the participants spent 1-4 hours a week, and the remaining 14% spent less than 1 hour a week using it. There were three participants who indicated they had limited experience in Palm OS, Pocket IE PDA, and WAP which were not related to Web e-learning. Thus, all participants had less than 1 hour's experience in Pocket IE and Opera browser through the training session provided. Of the 36 participants, 9 (25%) participants had only Internet Explorer browser experience, 16 (44.4%) participants had additional Netscape browser experience, 15 (41.6%) participants had additional Mozilla experience, and 4 (11.1%) participants had additional Opera PC browser experience.

The questionnaire also asked participants to select which interactive experiences they are familiar with, out of 20 possibilities. The survey showed that almost all (35 or 97%) of the participants were familiar with sending short messages (SMS). Most (69% or 25) of the participants were familiar with mobile phone games. Half the participants were familiar with using a camera phone. Less than half (44% or 16) of the participants were familiar with a multimedia messages service (MMS). The same number had interaction experience as a personal tutor. Only 25% (9) of the participants were familiar with mobile phone browsing using WAP and only 16% (6) of the participants were

Table 1. Abbreviations

Dimensions	Items	Items Meaning
Content D1	**CON-GR**	Granule
	CON-MU	Multimedia
	CON-OR	Organisation
	CON-PE	Pedagogy
Learner D2	**LER-EN**	Environment
	LER-HE	Help
	LER-MO	Motivation
	LER-PR	Preference
Device D3	**DEV-ER**	Ergonomic
	DEV-PR	Presentation
Communication D4	**COM-DE**	Delivery Mode
	COM-ST	Stability
	COM-RE	Response
Coordination D5	**COO-IN**	Interaction
	COO-NA	Navigation
	COO-PR	Presentation

familiar with PDA. A high proportion (75% or 25) of the participants was familiar with an e-learning system. This is because the university uses Blackboard as its e-learning system platform. Notice that the mobile phone and PDA responses regarding browsing experience differ from the interactive experience figures. This may be due to different interpretations from the participants. Nevertheless, it indicates that not many participants were experienced with PDA and mobile phone browsing for e-learning purposes.

EVALUATION

Based on the finding and useful comments in a pilot evaluation earlier, a full evaluation was conducted. In the full evaluation, e-learning system B has been improved on link navigations and includes multimedia with interaction during the adaptation process based on the comments from the pilot evaluation. The actual PDA used was IPAQ5450 and Pocket PC Internet explorer 2003 with real time Bluetooth connection. The actual mobile phone used was Nokia 6600 with Opera browser and Vodafone GPRS real time connection. An example of the e-learning module is shown in Figure 3. Figure 3(a) shows the text version with a PDA and Figure 3(b) shows an animated version.

Overall Learner Satisfaction Score

Table 2 depicts the mean overall learner satisfaction score with respect to different experiment scenarios organised by accessing devices and e-learning systems. This part of the analysis is to support evaluation objective 1 (EO1). Each individual cell in the table is an average over six participants on the overall learner satisfaction score taken from part four of the multiplatform e-learning system questionnaires. The overall learner satisfaction questions contain six items where the participants responded according to their learner experience after each completed task. It provides an aggregate score on interacting with the e-learning systems without getting into detailed measures. From Table 2, the following general observations can be made:

a. From the overall ranking and scores, participants seemed to prefer multiplatform e-learning system B rather than e-learning system A, which is a traditional e-learning system based on the Blackboard e-learning system. The implication could be that the multiplatform e-learning system was capable of adapting to the experiment situations

Figure 3a (left) Text content; b (right) Animation content-with permission from OSIM

and managed to provide better content as perceived by the participants than e-learning system A.

b. By comparing learner satisfaction scores according to accessing devices, participants seemed to accord higher learner satisfaction scores for PCs than PDAs and for PDAs than mobile phones, in that order, within each e-learning system. The ordering of the accessing devices may suggest that participants perceive that smaller mobile accessing devices will not be more satisfying than bigger desktop devices unless the adaptation process can modify such perception.

c. With an appropriate adaptation process, a mobile phone can achieve a better learner satisfaction score in multiplatform e-learning system B than a PDA device without adaptation in e-learning system A. Moreover, a PDA device with appropriate adaptation can achieve a better overall learner satisfaction score than a PC without adaptation. Therefore, it may imply that with the right adaptation process, a learner's pre-perception can be altered.

d. The best overall ranking is a PC with multiplatform e-learning system B. It achieved number one ranking and attained an average score of 6.7 out of 9. The worst ranking is a mobile phone with e-learning system A. It achieved an average score of 2.7 out of 9. Thus, mobile phone access with no appropriate adaptation is unlikely to be successful.

Gain in Multiplatform E-Learning System

In general, the overall learner satisfaction score for system B is higher than system A. To support evaluation objective 2 (EQ2), a system gain can be defined as:

$$Gain = \frac{LSS_B - LSS_A}{LSS_A}$$

To determine if the difference is statistically significant, ANOVA paired samples tests for repeated measures data were performed between system A and system B with the overall learner satisfaction score as the dependent variable and the e-learning system as the factor over three different accessing devices. Table 3 depicts the respective system gain factors. The results of the ANOVA test clearly indicate that mobile phone achieved the highest gain and is significant at 95% confidence level over all the evaluation scenarios. This result also

Table 2. Mean overall learner satisfaction score (LSS)

Scenarios	PC		PDA		Mobile	
	A	**B**	**A**	**B**	**A**	**B**
S1: Benchmark	7.0	7.3	5.6	6.0	2.8	5.9
S2: Off-line	6.9	7.5	6.3	7.0	2.1	5.2
S3: Interruption	5.7	6.3	4.9	6.1	2.7	5.2
S4: Pre-fetch	5.8	6.4	5.6	7.2	3.5	6.4
S5: Urgency	5.7	5.7	4.8	6.2	3.0	4.5
S6: Self Explore	7.5	7.5	5.0	6.4	2.2	6.4
Total Score	38.6	40.7	32.2	38.9	16.3	33.6
Mean score	6.43	6.78	5.37	6.48	2.72	5.60
Rank	3	1	5	2	6	4

provides a good indication that content adaptation for mobile phone access could achieve a very good return in the learner satisfaction score and increase the likelihood for adopting the multiplatform e-learning system. On the other hand, PC achieves a mere gain of 5% and is not significant at 95% confidence interval. This indicates that the adaptation process on the PC did not achieve as desired. One possible explanation is that the PC is the most familiar system and the learner is capable of adapting to the system. Another reason might be the scenarios evaluated did not trigger many significant differences between the two systems. This may suggest that the adaptation process may not be uniformly implemented across different devices and a more specific adaptation process may be required.

Degree of Influence and Relationships

One of the key research questions in this study is to examine and validate the degree of influence on overall learning satisfaction by each of the dimensions and sub-dimension factors identified in the multiplatform e-learning systems framework. This is the main purpose of evaluation objective 3 (EO3). To achieve this objective, Pearson's correlation analysis is used to establish the required statistics. The use of the correlation technique is recommended to avoid masking by strong factors where the factors had been identified from theories (Hair, Anderson, Tatham, & Black, 1995). Correlation analysis between all the factors and overall learner satisfaction produced a set of correlation coefficients and their significance levels. The larger the coefficient value, the stronger the relationship is between the factor and the overall learner satisfaction measure. Upon generating the correlation coefficients, each factor was ranked according to its value for ease of comparison. Subsequently, all the factors were averaged within the content dimension (D1), learner dimension (D2), device dimension (D3), communication dimension (D4), and coordination dimension (D5) respectively, to produce a composite indicator. Correlation analysis were performed independently for e-learning system A and multiplatform e-learning system B.

By examining the statistics that were generated in Table 4, several assessments can be made. First, all the correlation coefficients were positive and statistically significant at 0.001 levels except LER-EN ($p<0.05$). This implies that all the factors were positively influencing the overall learner satisfaction score. It also meant that the factors identified in the framework were all valid. Second, the degree of influence among the factors on the overall learner satisfaction score was not uniform. This meant that some factors can be more influential than others.

With reference to correlation statistics for e-learning system A and the multiplatform e-learning system in Table 4, the following observations can be made:

a. In e-learning system A, the maximum correlation coefficient was due to content organisation (CON-OR, 0.86) while the minimum correlation coefficient was due to learner environment (LER-EN, 0.05). The spread was about 0.81. The correlation

Table 3. System gain comparison

Device	System A Mean LSS	System B Mean LSS	Gain %	t-value	Df	Sig. difference (2-tailed)
PC	6.4	6.8	5	-1.928	35	0.062
PDA	5.4	6.5	20	-3.921	35	0.000
Mobile Phone	2.7	5.6	103	-19.076	35	0.000

coefficient for multiplatform e-learning system B had a range from 0.76 (CON-OR) to 0.18 (LER-EN). The spread was about 0.58, which is lower than e-learning system A. A wider spread in a system from observation during the experiment is normally associated with more variations and less stability and less consistency in providing satisfying user experience. Thus, participants may have perceived multiplatform e-learning system B to be more stable than e-learning system A because of its adaptive capability. The consequence of a stable e-learning system is that it can facilitate better opportunities for learning.

b. The average correlation coefficients by dimension in multiplatform e-learning system B had a narrower spread than in e-learning system A. The implication is that all dimensions are treated equally. An equalisation effect has occurred.

c. The top five factors perceived most influential for e-learning system A were organisation (CON-OR), delivery (COM-DE), navigation (COO-NA), ergonomics (DEV-ER), and response (COM-RE). The top five factors perceived most influential for multiplatform e-learning system B were organisation (CON-OR), delivery (COM-DE), navigation (COO-NA), interaction (COO-IN), and pedagogy (CON-PE). The pedagogy factor (CON-PE) in the content dimension and interaction (COO-IN) in the coordination dimension emerged and replaced perceived response (COM-RE) and ergonomics (DEV-ER) in e-learning system A as the influential factors in e-learning system B. As e-learning system B was perceived to be more stable than e-learning system A, several negative effects due to lack of adaptive capability would have been overcome. In e-learning system A, participants had expressed difficulty in maneuvering, especially using mobile devices. Moreover, content could be missing due to lack of adaptive capability. Once these obstacles had been overcome in multiplatform e-learning system B, participants could address higher e-learning issues such pedagogy and interactions.

d. In e-learning system A, the dimensions that had a stronger influence on overall learner satisfaction were the coordination dimension (D5), device dimension (D3), communication dimension (D4), content dimension (D1), and learner dimension (D2), in that order. The learner dimension (D2) remained the least influential dimension. In multiplatform e-learning system B, the dimensions that had stronger influence on overall learner satisfaction were the coordination dimension (D5), content dimension (D1), device dimension (D3), communication dimension (D4), and learner dimension (D2), in that order. This order is distinctly different from the order of e-learning system A. The significant difference was that the content dimension (D1) was the second most influential factor and the device dimension (D3) was the third most influential. Though the learner dimension (D2) remained the least influential dimension, the average correlation coefficient improved significantly. The reason for the content dimension to be more significant in system B is probably that adaptive capability had allowed participants to consider higher learning issues which are not possible in e-learning system A, where participants were busy dealing with device issues.

Grouping Behaviour in Factors

So far the factors within the multiplatform e-learning systems framework were validated and shown to be statistically significant in influencing overall learner satisfaction through correlation analysis. This part of the analysis is to support evaluation objective 4 (EQ4). By comparing the correlation coefficients between multiplatform e-learning system B and e-learning system A, one noticed that the correlation coefficients had been equalised in multiplatform e-learning system B. Equalisation means that the coefficient values are closer to each other and the spread is smaller. This effect may imply that these factors were becoming equally influential and important, as perceived by learners. This effect can only be noticed if two systems were used to compare. Thus, a comparative study can provide such insight while studying a single system alone cannot, unless the experiment is repeated after enhancement. Using a simple trend analysis method, Table 4 showed a trend of the differences in correlation coefficients between multiplatform e-learning system B and e-learning system A. A positive value indicates that the correlation coefficient is higher in e-learning system B and a negative value indicates otherwise. With reference to Table 4, the factors of environment (LER-EN), help (LER-HE), motivation (LER-MO), and preference (LER-PR) in the learner dimension (D2) and granularity (CON-GR) and pedagogy (CON-PE) in the content dimension (D1) improved their correlation coefficients. The improvement in the values indicates that learners had the opportunity to assess these factors as influential. This could happen provided the ergonomic obstacles in e-learning system A have been resolved or reached a threshold level of satisfaction. Thus the factors LER-EN, LER-HE, LER-MO, LER-PR, CON-GR, and CON-PE are termed secondary factors as they appeared more influential and relevant once the primary factors reached a threshold. Conversely, the factors in D3, D4, and D5 all have lower correlation coefficients in e-learning system B than in e-learning system A. These are the dominant factors that influence overall learner satisfaction in the non adaptive e-learning system A. These dominant factors are termed primary factors as they provided the main influences. Primary factors are indicated as a negative sign and secondary factors are indicated as a positive sign in the last column of Table 4. A high positive value indicates a strong secondary factor while a high negative value indicates a strong primary factor. In multiplatform e-learning system B, correlation coefficients of primary factors decreased while secondary factors increased.

Thus far, two groups of factors among the correlation coefficients have been identified. Primary factors were dominated by D3, D4, and D5 while secondary factors were dominated by D1 and D2. Designers can take advantage of the features by observing the emergence of the equalisation effect and secondary factors as an indication of a stable and well adapted e-learning system. This is likely to happen when the multiplatform e-learning system overcomes the threshold dominated by primary influencing factors.

Qualitative Data

Apart from quantitative data, qualitative data were collected through the interview and observation of the participant. When asked what they dislike about e-learning system A, one participant commented as follows:

I dislike that the e-learning system is often confusing and hard to navigate using the tools on the side bar, it has quite frequent disruptions and has either not enough information to get you where you want to go in a hurry, or it takes a while to get there through many steps.

When asked what they dislike about e-learning system B, one participant commented as follows:

The only thing that I was unsure of was that it wasn't immediately clear if one had answered all the questions correctly at the end as no "final score" was given. The style of testing meant that there could technically be no competition aspect to the test. While you would know straight away if you had something wrong, you would still be able to "pass" simply by trying each answer until you got the right one.

When asked what they like about e-learning system A, one participant commented as follows:

I like the learning style, it allows you to learn at your own pace and go over information as necessary.

When asked what they like about e-learning system B, one participant commented as follows:

I liked the look very much; it was much simpler and clearer to read than the BB site. While looking at the information on BlackBoard I found my *eyes skimming over information a bit, it was hard to concentrate on all the words. I would probably prefer a more rounded font as it's easier to look at. But the layout was good, the home button is helpful and it took me where I expected it to take me.*

Other participants have similar comments and, in general, find e-learning system B much more satisfying then e-learning system A. Participants also commented negatively on the navigation aspect of e-learning system A with mobile phone access. These comments coincide with previous findings.

With respect to PDA, participants find the device interface not so easy to use. A participant commented about the "zooming" and "show all" features of the PDA while exploring the embedded flash animation as follows:

While I liked both the flash demonstrations in this, I found it hard to use the pen to change the settings on them and had to resort to zooming in up close to make the pen move the animations, then move the whole screen around to see the result. It was hard to navigate and I couldn't see the entire effect too clearly always…

This comment could imply that while flash animation is included as part of the learning content to improve learning, the inability to use it satisfactorily might influence the learner satisfaction, subsequently resulting in a negative experience overall.

CONCLUSION

This chapter has presented a framework for the multiplatform e-learning system and its evaluation. Learner interactions with the systems were perturbed through six scenario evaluations. Factors and its relationship in influencing overall learner satisfaction within the multiplatform e-learning system framework were explored and confirmed. The full evaluation and statistical test confirmed that the multiplatform e-learning system achieved a higher gain with mobile devices. Overall, the mean overall learner satisfaction score is higher for the multiplatform e-learning system than the traditional e-learning system. The correlation analysis confirmed the significance of the factors influencing overall learner satisfaction. The most highly influencing factors were identified to be content organization (CON-OR), delivery mode (COM-DE), navigation (COO-NA), and interaction (COO-IN) in system B. Further analysis revealed two different groups of factors. Primary factors dominate in the non adaptive e-learning system and secondary factors emerged in the adaptive e-learning system due to the equalisation effect. The significance and relationship of the factors and correlation coefficients allow the designer to set criteria and further improve the adaptive e-learning system to achieve a better understanding and adoption of the multiplatform e-learning system.

The evaluation and results come with some limitations. First, the evaluation was conducted while learners had short term use of the e-learning systems. Though e-learning system A represents a familiar site with longer term use, most learners are somehow not familiar with its content. Thus, further evaluation would need to determine the long term effect when learners are familiar with the e-learning system and its content. Second, while the scenarios evaluation intends to perturb the interactions between the learner and the system as much as possible, there is no guarantee that all the factors will be equally perturbed. It is also unlikely that one can control only a particular factor at one time to capture the effect independently, as these factors act together. Third, the pedagogy of e-learning systems intends to expose the learner first with theory and principles then by multimedia simulation and followed by exercise. During the exercise, learners were encouraged to revisit the content and interactive multimedia. The evaluation was based on such pedagogy.

With respect to future work, one area that can benefit future research with respect to theory is to examine the relationship between multiplatform e-learning systems and multiple representations theory. Another area with respect to implementation is to explore other technologies to implement

*Table 4. Dimension ranking, factors ranking, and correlation coefficients *P<0.05, all others p<0.001*

			System A				System B					
Dimensions	**Factor**	**Factors Meaning**	**Factor Correlation Coefficient**	**Factor Rank**	**Average Coefficient by Dimension**	**Dimension Rank**	**Factor Correlation Coefficient**	**Factor Rank**	**Average Coefficient by Dimension**	**Dimension Rank**	**Changes in Correlation Coefficient**	**Primary and secondary factor**
Content D1	**CON-GR**	Granule	0.50	13	0.67	4	0.61	10	0.66	2	0.11	+
	CON-MU	Multimedia	0.74	6			0.58	11			-0.16	-
	CON-OR	Organisation	0.86	1			0.76	1			-0.09	-
	CON-PE	Pedagogy	0.58	12			0.67	5			0.09	+
Learner D2	**LER-EN**	Environment	0.05*	16	0.33	5	0.18*	16	0.44	5	0.13	+
	LER-HE	Help	0.27	15			0.45	15			0.18	+
	LER-MO	Motivation	0.39	14			0.51	14			0.12	+
	LER-PR	Preference	0.60	11			0.62	9			0.02	+
Device D3	**DEV-ER**	Ergonomic	0.80	4	0.75	2	0.65	8	0.65	3	-0.15	-
	DEV-PR	Presentation	0.69	9			0.65	6			-0.04	-
Communication D4	**COM-DE**	Delivery Mode	0.84	2	0.74	3	0.73	2	0.62	4	-0.11	-
	COM-ST	Stability	0.63	10			0.57	13			-0.06	-
	COM-RE	Response	0.74	5			0.57	12			-0.18	-
Coordination D5	**COO-IN**	Interaction	0.72	8	0.75	1	0.69	4	0.69	1	-0.02	-
	COO-NA	Navigation	0.81	3			0.71	3			-0.10	-
	COO-PR	Presentation	0.72	7			0.65	7			-0.07	-

multiplatform e-learning systems. For example, implementations using autonomous agents and Web services may generate a more robust multiplatform e-learning system. With respect to platform, there is a need to explore other platforms with very low similarity content factors. One such example is to explore short message service (SMS) and multimedia messages service (MMS) as additional deliver channels. Another possibility is to explore iPod/iPhone or similar devices as alternative delivery channels. It will be interesting to see if the popularity, trendiness, and ease of use factors of such devices could override its limitation to deliver satisfying learning.

REFERENCES

Angin, O., Campbell, A. T., Kounavis, M. E., & Liao, R.R-F. (1998). The mobiware toolkit: Programmable support for adaptive mobile networking. *IEEE Personal Communications*, *5*(4), 32–43.

Bharadvaj, H., Joshi, A., & Auephanwiriyakul, S. (1998, October). An active transcoding proxy to support mobile Web access. In *Proceedings of the 17th IEEE Symposium on Reliable Distributed Systems.*

Bickmore, T. W., & Shilit, B. N. (1997). Digestor: Device-independent access to the World-Wide-Web. In *Proceedings of the 6th WWW Conference*.

Blackboard. (2006). *Academic suite*. Retrieved November 25, 2006, from http://www.blackboard.com/us/index.Bb

Boyle, C., & Encarnacionj, A. O. (1994). MetaDoc: An adaptive hypertext reading system. *User Models and User Adapted Interaction (UMUAI)*, *4*, 1-19.

Brusilovsky, P., & Pesin, L. (1994). LSLS-Tutor: An adaptive hypertext learning environment. In H. Ueono & V. Stefanuk (Eds.), *Proceedings of Japanese-CIS Symposium on knowledge-based software engineering (JCKBSE'94)*. Tokyo: EIC.

Buyukkokten, O., Garcia-Molina, H., Paepcke, A., & Winograd, T. (2000, April). Power browser: Efficient Web browsing for PDAs. In *Proceedings of the CHI2000*. The Hague.

Chen, J., Yang, Y., & Zhang, H. (2000, August). An adaptive Web content delivery system. In *Proceedings of the International Conference on Adaptive Hypermedia and Adaptive Web-based Systems (AH2000)* (pp. 28–30). Italy.

Chen, J., Zhou, B., Shi, J., Zhang, H., & Wu, Q. (2001, May). Functional-based object model towards Web site adaptation. In *Proceedings of the WWW10* (pp. 1–5). Hong Kong. http://www10.org/ cdrom/papers/296/

Chin, J. P., Diehl, V. A., & Norman, K. L. (1988). Development of an instrument measuring user satisfaction of the human-computer interface. In *Proceedings of the ACM CHI '88* (pp. 213-218). Washington, DC:

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, *13*(3), 319-340.

Davis, F. D. (1993). User acceptance of information technology: System characteristics, user perceptions and behavioral impacts. *International Journal of Man-Machine Studies*, *38*(3), 475-487.

Feyer, T., Kao, O., Schewe, K-D., & Thalheim, B. (2000). Design of data intensive Web based information services. In *Proceedings of the 1st International Conference on Web Information Systems Engineering.* Hong Kong, China.

Finkelstein, A., Savigni, A., Kappel, G., Retschitzegger, W., Kimmerstorfer, E., Schwinger, W., Hofer, T., Pröll, B., & Feichtner, C. (2002). Ubiquitous Web application development: A framework for understanding. In *Proceedings of the 6th World Multiconference on Systemics,Cybernetics and Informatics (SCI 2002)* (pp.431–438). Orlando, Florida.

Fox, A., Goldberg, I., Gribble, S. D., Lee, D. C., Polito, A., & Brewer, E. A. (1998, September). Experience with top gun wingman, a proxy-based graphical Web browser for the USR PalmPilot. In

Proceedings of the IFIP International Conference on Distributed Systems Platforms and Open Distributed Processing (Middleware '98). Lake District, UK.

Fox, A., Gribble, S. D., Chawathe, Y., & Brewer, E. A. (1998, August). Adapting to network and client variation using active proxies: Lessons and perspectives. *IEEE Personal Communication, 5*(4),10–19.

Goh, T., & Kinshuk, (2004, September 20-24). *An adaptation framework for Web based learning system*. Paper presented at the Knowledge-Based Intelligent Information and Engineering Systems: 8th International Conference (KES 2004). Wellington, New Zealand.

Goldfarb, C. F., & Prescod, P. (1998). *The XML handbook*. New Jersey: Prentice Hall.

Hair, J. F., Anderson, R. E., Tatham, R. L., & Black, W. C. (1995). *Multivariate data analysis*. Englewood Cliffs, NJ: Prentice Hall.

Hook, K. (1998). Evaluating the utility and usability of an adaptive hypermedia system. *Journal of Knowledge-Based Systems, 10*, 311-319.

Hori, M., Kondoh, G., Ono, K., Hirose, S., & Singhal, S. (2000, May 15-19). Annotation-based Web content transcoding. In *Proceedings of the Ninth International World Wide Web Conference*. Amsterdam, The Netherlands.

Intel. (1998). Intel QuickWeb. Retrieved November 10, 2002, from http://www.intel.com/pressroom/archive/releases/IN011998.HTM

Johnson, T. R., Zhang, J., Tang, Z., Johnson, C. M., & Turley, J. P. (2004). Assessing informatics students' satisfaction with a Web-based courseware system. *International Journal of Medical Informatics, 73*, 181-187.

Kappel, G., Retschitzegger, W., Kimmerstorfer, E., Proll, B., Schwinger, W., & Hofer, T. (2002). Towards a generic customisation model for ubiquitous Web applications. In *Proceedings of the 2nd International Workshop on Web Oriented Software Technology* (pp. 79–104) Malaga

Lewis, J. (2002). Psychometric evaluation of the PSSUQ using data from five years of usability studies. *International Journal of Human-Computer Interaction, 14*(3&4), 463-488.

Lin, H., Choong, Y.-Y., & Salvendy, G. (1997). A proposed index of usability: A method for comparing the relative usability of different software systems. *Behaviour & Information Technology, 16*(4), 267-278.

Ma, W-Y., Bedner, I., Chang, G., Kuchinsky, A., & Zhang, H. J. (2000). A framework for adaptive content delivery in heterogeneous network environments. HP Lab http://www.cooltown.hp.com/dev/wpapers/adcon/MMCN2000.asp

Meyer, B. (1994). Adaptive performance support: User acceptance of a self-adapting system. In *Proceedings of the 4th International Conference on UM*. Hyannis, MA.

Segall, N., Doolen T. L., & Porter, J. D. (2005). A usability comparison of PDA-based quizzes and paper-and-pencil quizzes. *Computers and Education, 45*(4), 417-432.

Smith, J., Mohan, R., & Li, C. (1999). *Scalable multimedia delivery for pervasive computing*. ACM Multimedia.

Tretiakov, A., & Kinshuk, (2004, August 30-September 1). A unified approach to mobile adaptation of educational content. In Kinshuk, C.-K. Looi, E. Sutinen, D. Sampson, I. Aedo, L. Uden, & E. Kähkönen (Eds.), *Proceedings of the 4th IEEE International Conference on Advanced learning Technologies 2004*. Joensuu, Finland (pp. 101-105). Los Alamitos, CA: IEEE Computer Society (ISBN 0-7695-2181-9).

UWA. (2002). The UWA approach to modeling ubiquitous Web applications. In *Proceedings of the IST Mobile and Wireless Telecommunications Summit 2002*. Thessaloniki, Greece.

W3C. (2001). *Device independence principles*. Retrieved September 19, 2005, from http://www.w3.org/TR/2001/WD-di-princ-20010918/

Weibelzahl, S. (2003). *Evaluation of adaptive systems*. Retrieved November 25, 2005, from http://ubdok.uni-trier.de/diss/diss11/20030428/20030428.htm

Zachman, J. A. (1987). A framework for information systems architecture. *IBM Systems Journal*, *26*(3), IBM Publication G321-5298.

KEY TERMS

Content Adaptation: Content adaptation is the ability of a system to adapt content to suit any accessing devices.

Correlation Analysis: Correlation analysis is a statistical method to measure the strength of the relationship between two variables. The correlation coefficient cannot be greater than 1 or less than -1.

E-Learning: E-learning refers to any type of learning situation when instruction content is delivered electronically via the Internet.

Learner Satisfaction: Learner satisfaction is aggregate feelings or affective responses to distinguishable factors while interacting with the e-learning system.

Mobile Learning: M-learning is the delivery of learning content to students through the use of mobile or portable technology.

Multiplatform E-Learning System: A multiplatform e-learning system is an e-learning system that can deliver learning content to different accessing devices such as PCs, PDAs, and mobile phones.

User Experience: User experience is defined as a set of material rendered by a user agent which may be perceived by a user and with which interaction may be possible.

About the Contributors

Joanna Lumsden is a research officer with the National Research Council of Canada's (NRC) Institute for Information Technology. Prior to joining the NRC, Joanna worked as a research assistant in the Computing Science Department at the University of Glasgow, U.K. where she attained both her undergraduate Software Engineering Honours Degree and her PhD in human computer interaction. Joanna is also an adjunct professor at the University of New Brunswick in Fredericton, where she teaches graduate courses and supervises a number of graduate students. Joanna is the lab manager for the NRC's Mobile Human Computer Interaction Lab—a facility dedicated to investigating mobile interaction design and evaluation.

Matt Jones is a senior lecturer and is helping to set up the Future Interaction Technology Lab at Swansea University. He has worked on mobile interaction issues for the past ten years and has published a large number of articles in this area. He is the co-author of *Mobile Interaction Design*, John Wiley & Sons (2006). He has had many collaborations and interactions with handset and service developers including Orange, Reuters, BT Cellnet, Nokia and Adaptive Info; and has one mobile patent pending. He is an editor of the *International Journal of Personal and Ubiquitous Computing* and on the steering committee for the Mobile Human Computer Interaction conference series. He is married with three mobile, small children; when he's not working he enjoys moving quickly on a bike while listening to music and the occasional podcast.

* * *

Julio Abascal (BSc in physics, Universidad de Navarra, 1978; PhD in informatics, University of the Basque Country-Euskal Herriko Unibertsitatea, UPV-EHU, 1987) is a professor at the Computer Architecture and Technology Department of the UPV-EHU (Spain), where he has been working since 1981. In 1985 he co-founded the Laboratory of Human-Computer Interaction for Special Needs where he has led several R&D projects. His research activity is focused on the application of human-computer interaction methods and techniques to assistive technology, including the design of ubiquitous, adaptive, and accessible user interfaces, and human-robot interfaces for smart wheelchairs. In addition, he leads a research group aiming to develop methods and tools to enhance physical and cognitive accessibility to the web. He has been Spain's representative on the IFIP Technical Committee 13 on "Human-Computer Interaction" since 1991, and is the former and founder chairman (1993) of IFIP WG 13.3 "Human-Computer Interaction and Disability." From 1990 he served as an adviser, reviewer, and evaluator for diverse EU research programmes (TIDE, TAP, IST…), and he is currently a member of the Management Committee of COST 219ter "Accessibility for All to Services and Terminals for Next Generation Networks."

Gregory D. Abowd is an associate professor in the College of Computing and GVU Center at the Georgia Institute of Technology, and director of the Aware Home Research Initiative. His research focuses on an application-driven approach to ubiquitous computing concerning both HCI and software engineering issues. He received a PhD in computation from the University of Oxford. He's a member of the IEEE Computer Society and the ACM.

Thomas Alexander heads the research group on 3D Visualization and Interaction at the FGAN-Research Institute for Communication, Information Processing, and Ergonomics (FKIE) in Wachtberg, Germany. He obtained the Diplom-Ingenieur (MS) and Dr.-Ing. (PhD) degree in 1994 and 2002, respectively. Since 1994 he has been working at FKIE, where he has been responsible for research activities and various projects in the area of digital human models and human-computer interaction (HCI). In 2004-05 he worked as a visiting scientist at the NASA Ames Research Center, CA. Since 1998 he has been German national representative and chairs several international committees on virtual environments.

Over the past three decades, **Rebecca Allen** has been recognized internationally for her ground-breaking work in media art and her research in 3D computer graphics, human motion simulation, artificial life, augmented reality and intimate interfaces. Allen is professor at UCLA Department of Design Media Arts and was founding chair of her department. Previous positions include: senior research scientist and director of the Liminal Devices research group at MIT Media Lab Europe; founding director of the Intel funded research project called Emergence; creative director and 3D visionary for video game company, Virgin Interactive; senior researcher at the world-renowned NYIT Computer Graphics Laboratory and member of the MIT Architecture Machine Group.

Nikolaos Avouris was born in Zakynthos, Greece (1956). He has a degree in electrical engineering from NTUA Greece (1979) and an MSc (1981) and PhD (1983) from the University of Manchester UMIST, UK. He served in various research positions in the UK, Italy, and Greece and then joined the University of Patras, as associate professor (1994-2001) and full professor of software engineering and human-computer interaction (2001-present). He is founder and head of the Human-Computer Interaction Group. His main interests are related to design and evaluation of interactive systems, usability engineering, collaboration technology, context-aware computing systems, and analysis and evaluation of collaborative activities.

Chris Baber received his PhD in human factors of speech technology from Aston University in 1990 before taking up a post at The University of Birmingham, where he is now reader in interactive systems design. His research focuses on human interaction with technology and he is particularly interested in how an understanding of everyday activities can be used to better inform the study of novel technologies.

Cédric Bach is a usability specialist of human virtual environment interactions (HVEI). He conducted his PhD in psychology at the French National Institute for Research in Computer Science and Control (INRIA) until 2004. Currently, he's a postdoctoral researcher in the Research Institute in Computer Science of Toulouse – France (IRIT). His research domain is methods for HVEI ergonomics dedicated to evaluation and user centred design.

Young Kyun Baek has been teaching at Korea National University of Education since 1991. Currently, Dr. Baek's research interest is on designing educational games and simulation, and using mobile devices in education. He has recently published two books entitled *Understanding & Applying Game Based Learning in Classroom* and *Understanding Edutainment* in Korea.

Rafael "Tico" Ballagas is a computer science doctoral candidate and research assistant at RWTH Aachen University in Aachen, Germany. His research interests include post-desktop user interfaces (including mobile phones), physical toolkits, and applying iterative design techniques to ubiquitous computing applications. He developed the iStuff toolkit to rapidly prototype new interactions in ubiquitous computing. He has an MS in electrical engineering from Stanford University and a BS in electrical engineering from Georgia Institute of Technology.

Susanne Bay is a psychologist with a special interest in cognitive ergonomics of small screen and medical devices. The work presented in the current chapter was carried out during her time as a PhD student at the Department of Psychology of RWTH Aachen University, Germany. Currently she is working as a human factors specialist.

Benjamin B. Bederson is an associate professor of computer science at the Human-Computer Interaction Lab at the Institute for Advanced Computer Studies at the University of Maryland, College Park. His work is on mobile device interfaces, information visualization, interaction strategies, digital libraries, and accessibility issues such as voting system usability.

Francesco Bellotti is a researcher at the ELIOS laboratory of the University of Genoa. He is involved in research projects concerning mobile computing, human-computer interaction and multimedia systems in the automotive environment. He received a PhD in electrical engineering from the University of Genoa. Professor Bellotti has authored over 50 papers in international journals and conferences.

Regina Bernhaupt is currently working as assistant professor at the HCI Unit of the ICT&S-Center, working on her habilitation in the area of usability evaluation methods. She holds a masters degree in psychology and in computer science from the Salzburg University. In 2002 she finished her technical dissertation in computer science in the field of intelligent systems (time coded artificial neural networks). She teaches programming courses, user interface techniques and design and human-computer interaction at the Salzburg University and the applied university of Salzburg. She is leading several projects in the area of home entertainment (interactive TV, games, new ways of entertainment) and is responsible for new forms of usability and user experience evaluation in various contexts like mobile interfaces and ambient technologies.

Riccardo Berta is a research consultant at the Department of Biophysical and Electronic Engineering of the University of Genova. He received his degree (MS) in electronic engineering from the University of Genova in 1999, with a thesis on project and development of an optimizing environment for Java bytecode. His current research interests include design, implementation and evaluation of innovative modalities of human-computer interaction for mobile devices. Dr. Berta has authored over 30 papers in international journals and conferences.

Enrico Bertini is a post-doc researcher at University of Fribourg, Switzerland. He holds a PhD in computer engineering from University of Rome "La Sapienza". His research interests span across the whole spectrum of human-computer interaction. In recent years, he has explored issues related to usability in mobile computing, devising novel methodologies for evaluation, and adaptation/personalization techniques. Information Visualization is also one of his main research interests.

Jason Black has a PhD in computer science from Florida State University, an MS degree in computer science from Georgia Institute of Technology, and a BS degree in computer information systems from Florida A&M University. His research interests include mobile computing, collaborative mobile learning (m-learning), user interface design for mobile devices, ambient or ubiquitous computing, and educational technology. He also has a particular interest in developing learning applications for personal digital assistants (PDAs) with a focus outside of math and science, such as reading comprehension. Dr. Black has been as assistant professor at FAMU for seven years.

Ann Blandford is professor of human-computer interaction and director of UCL Interaction Centre. There, she leads a team of 20 academics and researchers studying various aspects of HCI. Her research focuses on seeking a better understanding of interactive systems from a user perspective, and using that understanding to support the design and evaluation of systems. Her research activities include investigating the use of systems in context, and the effects of system design on user experience.

Borja Bonail works at the Laboratory of Human-Computer Interaction for Special Needs of the University of the Basque Country-Euskal Herriko Unibertsitatea, UPV-EHU (Spain) with a Pre-doctoral Fellowship granted by the Basque Government. He holds a BSc in informatics engineering (2004) from the UPV-EHU. His main research areas are the development of methodologies for human-robot interaction in shared-control mode, and the development of intelligent robotics controllers to support such interaction types, both focused on wheelchair users.

Jan Borchers is full professor of computer science and head of the Media Computing Group at RWTH Aachen University. With his research group, he explores the field of human-computer interaction, with a particular interest in new post-desktop user interfaces for smart environments, ubiquitous computing, interactive exhibits, and time-based media such as audio and video. Before joining RWTH, he worked as assistant professor of computer science at Stanford University for two years, and briefly at ETH Zurich. He received his PhD in computer science from Darmstadt University of Technology in 2000, and is a member of ACM, SIGCHI, BayCHI, and GI.

Mike Bradley is senior lecturer in Product Design and Engineering Group at Middlesex University, London, England. He has a BEng in mechanical engineering from Birmingham University, England and an MSc in ergonomics from Loughborough University, England. He was head of human factors at Ford of Europe and his research interests are in interaction design, inclusive design, and human-centred intelligent vehicles.

Stephen Brewster is a professor of human-computer interaction in the Department of Computing Science at the University of Glasgow, UK. Brewster joined Glasgow in 1995, before that working as an EU-funded ERCIM fellow in Finland and Norway. Brewster's work is in the area of multimodal human-computer interaction (HCI), using multiple sensory modalities to make human-computer interaction more effective. He has directed his research towards sound, touch, and smell. His work on sound has focused around the design and use of Earcons, or structured non-speech sounds. His research in this area is now on the use of spatial sound in mobile computing, where the lack of screen space means audio plays an important role. Brewster has also worked on gestural interfaces for mobile devices. One other strand of his work is in haptic (touch-based) interaction, here focusing on applications for visually-impaired people and medical simulation. Brewster is currently an EPSRC Advanced Research Fellow. The aim of the fellowship is to understand how to construct tactile cues for user interfaces, particularly Tactons, or tactile icons as another form of output for mobile devices.

Xavier Briffault is a researcher in the Centre National de la Recherche Scientifique (CNRS) in France. He is a computer scientist with a primary interest in natural language processing. He has been involved in various research projects concerning computer supported collaborative work, knowledge representation, object oriented programming, and multimodal navigation aids for car drivers and blind pedestrians.

Earl Bryenton is a professional engineer and a seasoned veteran of business management with over 50 years of industry experience. He founded his own consulting engineering company in 1983 that later became BRYTECH Inc. Prior to this he managed corporate operations at Bell Northern Research having ascended through the ranks including technical development, systems engineering, and marketing. Earl graduated from the University of New Brunswick with a BSc in engineering and holds a current license as a consulting engineer.

Thorsten Büring is a final year PhD candidate in the Department of Computer Science at the University of Konstanz, Germany. He is a member of the Human-Computer Interaction Group headed by professor Harald Reiterer. His research is supported by the Deutsche Forschungsgemeinschaft and focuses on mobile HCI and information visualization. He holds a MSc in information technology from the University of Glasgow (2003) and a Diploma in electronic business design from the University of Arts, Berlin and the Institute of Electronic Business (2002).

Gary Burnett has 15 years experience in human factors research and development relating to advanced technology within road-based vehicles. He has been a lecturer in human factors and human-computer interaction at the University of Nottingham in the UK since July 2000. Previously, he worked as a research fellow at the HUSAT Research Institute at Loughborough University. His work addresses key safety, usability, and acceptability issues for a number of in-car systems, and he has worked on a number of large-scale collaborative projects within this area (funded by the EU and the UK government). He also acted as a consultant to many of the major car manufacturers and system suppliers (e.g., Honda, Ford/Jaguar, Toyota, Nissan, Alpine). He has published over 50 papers in peer-reviewed journals, conferences, and edited works.

Daniel Cagigas received his PhD degree in computer science from the University of the Basque Country-Euskal Herriko Unibertsitatea, UPV-EHU (Spain) in 2001, for his work on navigation systems for smart wheelchairs. He is now a lecturer at the Department of Computer Architecture and Technology and a member of the 'Robotics and Technical Aids Group' research group, both at the University of Seville (Spain). He lectures on fundamentals of computer science, computer architecture, and parallel architecture systems. His research interests include artificial intelligence, robotics, human-computer interaction, and technologies for rehabilitation.

Mikael Cankar, MSc, is currently a usability specialist at TeliaSonera, where he works to improve the user experience of mobile and Web services. He has recently begun PhD studies and has ambitions to improve the way user research findings are used in product development projects.

Tolga Capin is an assistant professor in computer engineering at Bilkent University, Turkey. He has been involved in the development of mobile graphics and user interface technologies since 2000; most recently contributing to the development of graphics engines as a principal scientist and research manager at Nokia Research Center. His current research interests are mobile graphics platforms, user interaction with graphics and images, and networked graphics platforms. He got his PhD degree in computer science in 1998.

Tiziana Catarci received her PhD in computer science from the University of Rome, where she is currently a full professor. She has published over 100 papers and 10 books in a variety of subjects comprising user interfaces for databases, 2D and 3D data visualization, adaptive interfaces, visual metaphors, usability testing, data quality, cooperative database systems, database integration, and Web access. Dr. Catarci is regularly in the programming committees of the main database and human-computer interaction conferences and is associate editor of *ACM SIGMOD Digital Symposium Collection (DiSC), VLDB Journal, World Wide Web Journal,* and *Journal of Data Semantics.*

Nian-Shing Chen is professor of the Department of Information Management, National Sun Yat-sen University, Taiwan. He is currently the chairman of the MIS Department, NSYSU and adjunct professor of the Griffith Institute for Higher Education, Griffith University, Australia. He has published over 160 research papers in international refereed journals, conference, and book chapters. His research areas include e-learning, knowledge management, and the use and development of online synchronous learning and wireless technologies to enhance learning.

Eleni Christopoulou is a PhD candidate in computer science at the Department of Computer Engineering & Informatics, Patras University, Greece. She received her Engineering Diploma and MSc in computer science and engineering, in 2001 and 2004, respectively, both from the same institution. During 1999-2005 she worked as an R&D Engineer for the Research Academic Computer Technology Institute. During 2004-2005 she taught at the Technological Educational Institute of Patras, Greece. Since 2006 she teaches at the Department of Informatics, Ionian University, Greece. Her current research interests include the use of ontologies and context-awareness in ubiquitous computing systems and semantics in p2p networks and social networks.

Martin Colbert graduated in psychology from the University of Stirling, UK, in 1985. From 1986 to 1998, he was a research associate at the Ergonomics & HCI Unit, University College London (UCL), UK, where he gained his PhD (1995). In 1998, he became senior lecturer at the School of Computing and Information Systems, Kingston University, UK, where he now teaches Human Computer Interaction and Java, and is course leader for MSc Software Engineering. He continues to research and publish in HCI, and conducts contract usability work. Martin likes his family and friends, mountains, bicycles, music and parties, preferably all together.

Jose L. Contreras-Vidal is an associate professor in kinesiology, neuroscience, and bioengineering at the University of Maryland, College Park. His research integrates behavioral, neuroimaging, and computational neuroscience methods to study the neural mechanisms and computational principles underlying human sensory-motor control and learning. His research also includes computer simulations of large-scale, biologically-plausible, neural networks of cognitive-motor systems.

Enrico Costanza is a doctoral student and research assistant at the Laboratoire de design et media at the Ecole Polytechnique Fédérale de Lausanne. He holds an MEng in electronics and communications engineering from the University of York, England, and an MS in media art and science from the Massachusetts Institute of Technology, USA. From 2003 to 2005 he was a research associate at the Media Lab Europe in Dublin, Ireland. Enrico is active in the areas of mobile HCI, visual marker recognition and its HCI applications, and tangible user interfaces.

David Coyle is a PhD candidate with the Department of Computer Science at Trinity College Dublin. His background is interdisciplinary, in areas including electronic engineering, multimedia, computer science, and human computer interaction. His current research is focused on developing user-centred design methodologies for the design and evaluation of adaptable technologies for talk-based mental health interventions. David is currently exploring the design and impact of therapeutic 3D computer games in adolescent mental health interventions.

Murray Crease is a research officer with the National Research Council of Canada based in Fredericton, New Brunswick. He received his BSc (1996) and PhD (2001) in computing science from the University of Glasgow, Scotland. Murray's research interests lie in the general area of human-computer interaction (mobile HCI). Murray is particularly interested in mobile HCI believing that mobile systems should not be limited to mobile devices such as PDAs and mobile phones but should also include ubiquitous computing and alternative interaction techniques such tangible and ambient interaction.

Andrew Crossan graduated with a BSc Hons at the Department of Computing Science at Glasgow University in 1999. He then completed his PhD at the same department before moving to the Hamilton Institute in 2003 where he worked as a research assistant studying continuous control interaction techniques with mobile devices. Since 2005 he has worked as a post-doctoral researcher in the Multimodal Interaction Group (MIG) in the University of Glasgow Computing Science Department. One major focus of his work has been on multimodal interaction, with the main applications being in virtual reality veterinary medical training systems and in developing accessible interfaces for visually impaired people.

Sara Dalzel-Job received an MA in psychology from the University of Glasgow in 2004. She is currently studying for an MSc in informatics at the University of Edinburgh (due to be completed in August 2007), to be followed by a PhD in the same department. Her current interest is human computer interaction, specifically the use of eye tracking technology as a diagnostic tool.

Nelly de Bonnefoy is a PhD student in a joint laboratory involving the IRIT laboratory and EADS. Her work focuses on operators in the context of aeronautical maintenance and the main goal is to "Bring the information they need, when they need it in the most user-friendly and intuitive ways". Three major axes characterise her approach: information filtering, presentation of the useful information, and interaction dynamics for a spatial augmented reality system.

Alessandro De Gloria is full professor of electronics at the University of Genoa. He is the leader of the ELIOS (Electronics for the Information Society) laboratory of the Department of Electronics and Biophysical Engineering at the University of Genoa. His research interests include virtual reality, mobile computing, human-computer interaction, and embedded system design. He has participated and directed several European research projects on these items. Professor De Gloria has authored over 100 papers in international journals and conferences. He is a member of the IEEE.

José Eustáquio Rangel de Queiroz received his electrical engineering degree from the Federal University of Paraíba, Brazil, in 1986 and the DSc degree in electrical engineering from the Federal University of Paraíba (UFPB), Brazil, in 2001. He was a senior engineer at the Brazilian National Institute for Space Research (Instituto Nacional de Pesquisas Espaciais - INPE), from 1986 to 1999, and a lecturer at the Computer Science Department at the Federal University of Campina Grande (UFCG), Brazil, since 2002. His research interests include human-computer interaction, digital image processing and GIS applications.

Li Deng received the bachelor's degree from the University of Science and Technology of China and received the PhD degree from the University of Wisconsin-Madison. In 1989, he joined the Department of Electrical and Computer Engineering, University of Waterloo, Ontario, Canada and became a full professor in 1996. In 1999, he joined Microsoft Research, Redmond, WA as a senior researcher, where he is currently a Principal Researcher. He is also an affiliate professor in the Department of Electrical Engineering at University of Washington, Seattle. He has published over 250 refereed papers in leading international conferences and journals, 12 book chapters, and has given keynotes, tutorials, and lectures worldwide. He has been granted over a dozen US or international patents. He authored two books in speech processing. He is fellow of the IEEE and fellow of the Acoustical Society of America.

Anind Dey is an assistant professor in the human-computer interaction Institute at Carnegie Mellon University. Anind conducts research at the intersection of ubiquitous computing and human-computer interaction. In particular, he has been performing research in context-aware computing for over a decade, looking at issues of end-user control and feedback, application development and tools for building context-aware applications, privacy and information overload. Before joining CMU, Anind was a senior researcher with Intel Research in Berkeley and an adjunct professor at the University of California, Berkeley. He received his PhD in 2000 in computer science from the Georgia Institute of Technology.

Alan Dix is professor in the Department of Computing, Lancaster University, UK. His interests include delays and temporal issues, interaction in mobile and ubiquitous systems, the use of formal methods in HCI, the design of cyberspace and the way information is transforming economics and society, bits of e-learning ... and just about everything.

Gavin Doherty is a lecturer in computer science at Trinity College Dublin and was principal investigator on the HEA "Virtually Healthy" project. Having originally qualified in computer science at TCD he obtained his doctorate in the Human Computer Interaction Group at the University of York. He later worked at CNR Istituto CNUCE, and the Rutherford Appleton Laboratory. His research interests include the application of technology to mental health interventions, and situations where standard approaches to user-centred design are not feasible. He is also interested in the design of visual information representations and user interactions with mobile technology.

Emmanuel Dubois achieved his PhD at the University of Grenoble in 2001 and is now lecturer at the University of Toulouse-IUT Tarbes-France. He is member of the Research Institute in Computer Science of Toulouse-France (IRIT). His main research domain is related to design methods and tools for augmented and mixed reality systems. His approach is based on the modeling of the different facets that form the user's interaction with such systems.

Mark Dunlop is a senior lecturer in the Department of Computing and Information Sciences at the University of Strathclyde in Glasgow, Scotland. He has investigated many aspects of mobile technology including visualisation of complex information, text entry and use of mobile devices to support lectures. He is a member of the international steering committee for MobileHCI and the editorial board of Personal and Ubiquitous Computing, and an associate editor of *Advances in Human Computer Interaction*. Prior to Strathclyde, Mark was a lecturer at The University of Glasgow and a senior researcher at Risø National Laboratory, Denmark.

Parisa Eslambolchilar received the BS degree in computer engineering from the University of Amirkabir (Tehran Polytechnic) in 1998, an MS degree in electrical engineering from the University of Tehran in 2001, and a PhD degree in computer science in 2006. She was a research fellow at the Hamilton Institute, National University of Ireland in Computer Science until 2007 and now works in the Future Interaction Technologies Laboratory at the University of Wales, UK.

Danilo de Sousa Ferreira is a masters student at the Federal University of Campina Grande, currently researching usability evaluation methods for mobile devices. He holds a Bachelor degree in computer science from the same university and has taken part in a fair number of usability evaluations of interactive systems. His main research interest is the usability evaluation of mobile devices.

Bob Fields is a lecturer in the School of Computing Science at Middlesex University, where he is a member of the Interaction Design Centre. His research is concerned with developing understandings of people's interactions with technology and with one another, and using such understandings in the design of new products, systems, and services. This research has investigated collaborative working and human-computer interaction in a number of domains, from aviation and air traffic control to mobile devices and domestic technology.

Georgios Fiotakis was born in Chania, Greece (1977). He holds a degree in electrical and computer engineering from the University of Patras, Greece. He is currently a PhD candidate at the same University pursuing doctoral research in usability engineering. His research interests lie in methods and tools for usability evaluation of mobile and collaborative applications. He is the main designer and developer of the ColAT and ActivityLens usability evaluation tools.

Clifton Forlines is a researcher at Mitsubishi Electric Research Laboratories. His research interests include: the design and evaluation of novel user interfaces, including digital video presentation; collaborative tabletops; multiuser, multidisplay workspaces; and hand-held projectors for augmented reality. Forlines has a masters of Human-Computer Interaction and a masters of Entertainment Technology from Carnegie Mellon University, and is pursuing a PhD in computer science at the University of Toronto.

Peter Fossick is a results-driven professional product designer with expertise in product, new media, and interaction design. He is a full-time academic in the UK. As a designer and as an academic he has worked for sustained periods in Europe, the USA, and SE Asia. He has wide ranging experience in developing high quality innovative, user centric products and services that have received critical acclaim and commercial success. As an academic he has been responsible for developing and leading a wide range of design courses at undergraduate and postgraduate level. As an active researcher he is interested in exploring service and interaction design issues, innovation strategies, e- manufacturing and mass-customisation in product design. He has a BA (Hons) from Manchester Polytechnic and an MSc in CAED from Strathclyde University.

Silvia Gabrielli got a PhD in cognitive sciences from the University of Padova (I) and since 1994 has been working in the field interaction design and usability evaluation of ICT. In recent years, her research focus has been on methodologies for usability and accessibility evaluation of mobile applications, as well as on the design and evaluation of educational environments. Silvia is currently working as a research fellow at HCI Lab (University of Udine, Italy).

Nestor Garay-Vitoria received a BSc degree in informatics (1990), an MSc degree in computer science and technology (1992,) and a PhD in computer science (2000) all from the University of the Basque Country-Euskal Herriko Unibertsitatea, UPV-EHU (Spain). His PhD dissertation received the University's Extraordinary Award in October 2002. Since 1991 he has been working in the Laboratory of Human-Computer Interaction for Special Needs, where he has collaborated on several research projects. These projects are mainly devoted to the alleviation of communication problems for people with severe motor disabilities. From 2004 onwards, he has led several projects related to affective computing. Currently he is an associate professor at the Computer Architecture and Technology Department of the UPV-EHU.

Luis Gardeazabal is an associate professor of the Department of Computer Architecture and Technology of the University of the Basque Country-Euskal Herriko Unibertsitatea, UPV-EHU (Spain). He was a co-founder of the Laboratory of Human-Computer Interaction for Special Needs (1985). He holds a BSc in physics (1981) from the University of Navarra. He also holds a BSc in engineering (1984) and a PhD in informatics for his work on Alternative and Augmentative Communications Systems for People with Disabilities (1999), both from the UPV-EHU. His main research area is the development of methodologies for special interaction devices for communication and mobility aids.

Florence Gaunet is a researcher in the Centre National de la Recherche Scientifique (CNRS), France. She is a cognitive psychologist and has been studying spatial representations in blind people for more than 10 years. She is experienced in human factors for human-computer interface design, and has an interest in spatial wayfinding by blind pedestrians and in the design of navigational aids for visually impaired persons.

David Gibson conducts research at the University of Vermont College of Engineering and Mathematics and is Project Investigator for the National Science Foundation ITEST project "The Global Challenge", a technology-based outreach program for high school students interested in solving global problems through science, technology, engineering, and mathematics. His recent edited book *Games and Simulations in Online Learning* presents research and development frameworks for the field.

Tiong Goh is currently a lecturer with the School of Information Management at Victoria University of Wellington, New Zealand. His research areas include mobile learning, mobile games for education, and mobile database applications development.

Philip Gray is a senior lecturer in the Computing Science Department at the University of Glasgow, Scotland. He has been actively engaged since 1983 in research into models, notations, software architectures and tools for user interfaces. Recently, he has focussed on the description and engineering of interaction techniques for mobile and ubiquitous systems. He is a director of Kelvin Connect Ltd, a small Scottish software company specialising in mobile information systems. He is also vice-chair of of IFIP Working Group 2.7/13.4 (User Interface Engineering) and the UK representative on IFIP Technical Committee 13 (Human Computer Interaction).

Cathal Gurrin is a senior postdoctoral researcher in the School of Computing, Dublin City University. He holds a PhD from DCU (2002) as well as a BSc degree, also from DCU. His PhD was in search engine algorithms and his current research is on developing, implementing, and evaluating efficient techniques for the management of all kinds of digital information including text, image, video, and even genomic data. He has published 44 journal and conference papers and posters, and in addition he serves on many international program committees annually.

Jonna Häkkilä works as a principal scientist at Nokia Research Center, Finland, doing research on mobile user experience. She received her PhD in 2007 from the University of Oulu, Finland. Her PhD research considered usability issues with context-aware mobile applications, and her current research activities include mobile context-awareness, haptic user interfaces, and mobile multimedia.

Dong-Han Ham is senior research fellow in the school of computing science at Middlesex University, United Kingdom. He received his BS in industrial engineering in 1989 from InHa University, and MS and PhD in industrial engineering in 1995 and 2001 respectively from Korea Advanced Institute of Science and Technology (KAIST). He was senior researcher in software & systems engineering at the Electronics and Telecommunications Research Institute (ETRI). His research interests include human-computer interaction joint with intelligent systems, human interaction with automation, software and systems engineering, human-computer interaction, and agent-oriented information systems.

Antonio Haro is a senior research engineer at D4D Technologies. He has worked in the areas of medical graphics, mobile computer vision, image and video based rendering, and augmented reality. He has been active in both computer vision and graphics research, on topics ranging from realistic skin microstructure detail capture and rendering using GPUs to real-time highly robust eye tracking on commodity hardware. He received his doctorate from Georgia Tech in 2003.

Bret Harsham joined MERL in 2001 to pursue interests in speech and multi-modal interfaces and speech-centric devices. Prior to joining MERL, Bret spent four years at Dragon Systems designing and implementing handheld and automotive speech products. Earlier, he was a principal architect of a Firewall and Virtual Private Network product. Bret has a BS in computer science and engineering from MIT.

Lois Wright Hawkes is a professor of computer science, associate dean of the College of Arts & Sciences at Florida State University, and Directory of the High-Performance Computing & Simulation Lab at the FAMU-FSU College of Engineering. Her PhD is from the University of London in electrical engineering (communications). Her research interests include the use of artificial intelligence and technology in education, fault tolerance, and high speed data transfer.

Gillian R. Hayes is a PhD candidate in the College of Computing and GVU Center at the Georgia Institute of Technology. Her research focuses on enabling capture of rich data in informal and unstructured settings and allowing access to that data. She received her BS in computer science and mathematics from Vanderbilt University. She's a member of the IEEE Computer Society and the ACM.

Paul C. Hébert, a critical care physician and clinical researcher, holds the rank of full professor in the Department of Medicine (Critical Care), University of Ottawa, with cross-appointments in the Departments of Epidemiology and Community Medicine, Anesthesiology and Surgery. Currently, he is vice-chair research of the Department of Medicine, University of Ottawa. Dr. Hébert has received several research awards, including "Researcher of the Year Award" from the Ottawa Health Research Institute, and the "Premier's Research Excellence Award" from the Ontario Ministry of Health (2002). With over 180 publications, his current interests include blood transfusion and conservation in critical care units.

Jeongyun Heo is senior researcher of the Mobile UI R&D Lab at LG Electronics Inc. She has received her MS in industrial engineering and BS in chemical engineering in 1991 and 1994, respectively, from the Korea Advanced Institute of Science and Technology (KAIST), where she is preparing her doctoral thesis. Her research has focused on user experience design which could link customer satisfaction to product design, and systematic approaches to achieving this. She is also interested in physical user interaction (PUI), usability issues related with the physical design, the methods of user research to catch the unmet needs considering usage context and culture difference, and usability of usability evaluation for quality improvement.

Steven Herbst has delivered competitively differentiated user experiences in domains ranging from mobile devices, to web applications, to medical equipment for over 15 years. Herbst leads the internationally-based Strategic Design Research team within Motorola's Consumer Experience Design organization. His team is responsible for revealing the underlying, often unarticulated, physical and digital design attributes that influence user satisfaction—and applying this knowledge to new products and platforms. Herbst has held positions at several other companies including Philips, AT&T, Lucent, and Prodigy. Herbst holds several user interface patents, and his research has influenced the designs of dozens of highly successful products.

Rune T. Høegh is a PhD student in the Human Computer Interaction Research Group of the Department of Computer Science at Aalborg University. His research focus is on how to improve the impact of usability activities in the software development process. His PhD project is called "Enhancing the Interplay Between Usability Evaluations and Software Design".

Giovanni Iachello is currently an associate at McKinsey & Co., focusing on IT and media strategy. He earned a PhD in computer science from Georgia Institute of Technology in 2006, with research on privacy and human-computer interaction. Previously, he worked as consultant in a Web services firm. He holds a Laurea from Padua University (Italy) and is member of the ACM, IEEE, IFIP WG9.6/11.7 and SIGCHI.

Samuel A. Inverso is a doctoral candidate in the Human Brain Dynamics Lab under the Visual Sciences Group in the Research School of Biological Sciences at the Australian National University in Canberra, Australia. He received a BS and MS degree in computer science from the Rochester Institute of Technology in Rochester, NY in 2001 and 2004, respectively, and was a research associate at the Media Lab Europe in Dublin, Ireland from 2003 to 2005. Sam's research interests thread computer science and cognitive neuroscience, and include brain-computer interfaces, subtle communication interfaces, and feedback in the cortical hierarchy.

Stephen Johnson graduated with an honours degree in electronic communications engineering from the University of Hull. He joined BT laboratories in 1989 to research statistical speech recognition algorithms and subsequently went on to develop innovative interactive voice demonstration systems. Stephen now works in BT's Mobility Research Centre where he is investigating context awareness as an enabler for mobile applications and services. Stephen is a chartered engineer and a member of the IET.

Gareth Jones is a senior lecturer in the School of Computing at Dublin City University. He obtained a BEng in electrical and electronic engineering in 1989 and a PhD examining the Application of Linguistic Models in Continuous Speech Recognition in 1994, from the University of Bristol. He was a member of the Speech, Vision and Robotics Group, Department of Engineering and Computer Laboratory, University of Cambridge, and lecturer at the University of Exeter. He was a research fellow at the Toshiba R&D Centre, Japan, at the Informedia Project at Carnegie Mellon University, and at the National Institute of Informatics, Japan.

Eija Kaasinen is head of the Media and Mobile Usability research team at VTT, Technical Research Centre of Finland. She has led field studies in several research projects concerning the development of mobile Internet and personal navigation services. Based on the integrated results of those studies, Eija Kaasinen has defined a framework to study user acceptance of mobile services. The framework, the technology acceptance model for nobile services, is currently being applied in several projects. In her recent research activities, Eija Kaasinen has been applying the model to study user acceptance of ubiquitous services.

Anne Kaikkonen is a usability specialist in Nokia Multimedia. She has over 15 years experience in using psychological knowledge in design and development; she has worked in Nokia for 9 years.

Titti Kallio, MA, serves as TeliaSonera's senior expert of design & usability. She has 20 years of experience in the usability area.

Achilles D. Kameas received his engineering diploma (1989) and his PhD in human-computer interaction (1995) from the Department of Computer Engineering and Informatics, Univ. of Patras, Hellas. Since 2003, he has been an assistant professor with the Hellenic Open University, where he teaches software design and engineering. He is also R&D manager with Research Academic Computer Technology Institute (CTI), where he is the head of Research Unit 3 (Applied Information Systems) and the founder of DAISy group (http://daisy.cti.gr). His research interests focus around the design of Ubiquitous Computing and Ambient Intelligence systems, Interaction models and the uses of Ontologies. On these topics he has co-edited 2 books, published over 40 papers, and co-organized several international conferences and workshops. He served as an elected member of the Disappearing Computer steering group and the Convivio network steering group. He is a member of ACM, IEEE CS and Hellenic AI Society.

Anu Kankainen, PhD, is a senior user experience specialist at Idean Research in Espoo, Finland. Anu is a psychologist with 13 years of experience in working together with engineers, designers and marketing people in several international and national projects focusing on mobile technologies.

Amy K. Karlson received her BA and MS in computer science from Johns Hopkins University, and holds an MS in computer science from the University of Maryland, College Park, where she is currently a PhD candidate with the Human-Computer Interaction Lab in the Department of Computer Science. Ms. Karlson's research focuses on developing interfaces and interaction methods to improve information access under the constraints of mobile computing.

Osamu Katai received his BE, ME and PhD degrees in 1969, 1971 and 1979, respectively, from the Faculty of Mechanical and Precision Engineering, Kyoto University. He has been with Kyoto University from 1974. Now, he is a professor at the Dept. of Systems Science, Graduate School of Informatics. From 1980 to 1981, he had been a visiting researcher at National Research Institute for Information Science and Automation (INRIA), France. His current research interests are on the harmonious symbiosis of artificial systems with natural systems including ecological design, ecological interface design, environmental spatial design, community design, etc.

Hiroshi Kawakami received his BE, ME and PhD degrees in 1987, 1989, and 1994, respectively, from the Faculty of Precision Engineering, Kyoto University. From 1989, he was an instructor at the Department of Information Technology, Faculty of Engineering, Okayama University. Since 1998, he has been an associate professor with the Department of Systems Science, Graduate School of Informatics, Kyoto University. His current research interests are on ecological and emergent system design, the co-operative synthesis method and knowledge engineering.

Aki Kekäläinen, BSc, has proudly tackled usability and user experience puzzles offered by a variety of devices and services for 8 years. He is currently working as a usability specialist at TeliaSonera.

Juha Kela is a VP of Finwe Ltd., formerly a research scientist at VTT, and a doctoral student in the University of Oulu's Department of Electrical and Information Engineering. His research interests include multimodal and mobile interaction. He has an MSc in information processing from the University of Oulu.

Julie A. Kientz is a PhD student in the College of Computing and GVU Center at the Georgia Institute of Technology. She is also an assistant director of the Aware Home Research Initiative. Her research focuses on collaborative access to individual experiences. She received her BS in computer science and engineering from the University of Toledo. She's a member of the IEEE Computer Society and the ACM.

Stephen Kimani is currently an academic and research member of Jomo Kenyatta University of Agriculture and Technology (Kenya) and is affiliated with the University of Rome "La Sapienza." He has been a post-doctoral researcher with the University of Rome "La Sapienza" (2004-2006). He holds a PhD in computer engineering (University of Rome "La Sapienza," Italy) and MSc in advanced computing (University of Bristol, UK). His main research interest is in human-computer interaction (HCI). In particular, as HCI relates to areas/aspects such as: user interfaces, usability, accessibility, visualization, visual information access, visual data mining, digital libraries, and ubiquitous computing.

Kinshuk is professor and director of the School of Computing and Information Systems at Athabasca University, Canada. He has been involved in large-scale research projects on adaptive educational environments and has published over 200 research papers in international refereed journals, conferences and book chapters. He is chair of the IEEE Technical Committee on Learning Technology and editor of the SSCI indexed *Journal of Educational Technology & Societ*y (ISSN 1436-4522).

Jesper Kjeldskov is associate professor in Computer Science at Aalborg University, Denmark. Jesper is doing research into "indexical" interaction design for context-aware mobile and pervasive computer systems and exploring new methods and techniques for studying the use of mobile and pervasive computer technologies in laboratory and field settings. He holds a PhD in computer science and engineering and a master of HCI and Sociology from Aalborg University.

Scott Klemmer is an assistant professor of computer science at Stanford University, where he co-directs the Human-Computer Interaction Group. He received a dual BA in art-semiotics and computer science from Brown University in 1999, and an MS and PhD in computer science from UC Berkeley in 2001 and 2004, respectively. His primary research focus is interaction techniques and design tools that enable integrated interactions with physical and digital artifacts and environments. He is a recipient of the UIST 2006 Best Paper Award and the 2006 Microsoft Research New Faculty Fellowship.

Vanja Kljajevic is a cognitive scientist who is currently a guest researcher at NewHeights Software, Ottawa, and a research associate with the Games & Media Group, Human Oriented Technology Lab, Carleton University, Ottawa. Her diverse interests include human-computer interaction, usability, serious games, and language processing in intact and impaired brains.

James F. Knight was awarded his PhD in the ergonomics of Wearable Computers at The University of Birmingham in 2002 and he has published widely on this topic. He currently works as a research fellow on projects funded by the UK Ministry of Defence and the European Union, at the University of Birmingham with interests in the effects of wearing and interacting with wearable computer devices on the human body.

Vassilis Komis was born in Viannos, Crete, Greece (1965). He holds a degree in mathematics from the University of Crete (1987), DEA (1989) and doctoral degrees (1993) in Teaching of Computer Science from the University of Paris 7 - Denis Diderot (Jussieu). He is currently associate professor in the Department of Educational Sciences and Early Childhood Education of the University of Patras. His publications and research interests concern the teaching of computer science, the pupils' representations in the new information technologies and the representations formed during the use of computers in class room, the integration of computers in education, and the conception and development of educational software.

Irina Kondratova is group leader of the People-Centred Technologies Group at the National Research Council of Canada Institute for Information Technology. She is a graduate of, and adjunct professor at, the University of New Brunswick's Engineering Faculty. Dr. Kondratova is a professional engineer with the Association of Professional Engineers for the Province of New Brunswick and has more than 25 years of research, business, and consulting experience. Dr. Kondratova does research work in the area of voice and multimodal technologies that allow the use of speech, along with traditional keyboard or stylus, to interact with computers. These technologies improve usability of mobile devices and allow for better data gathering in the field, or in emergency situations.

Siu Cheung Kong is an associate professor in the Department of Mathematics, Science, Social Sciences and Technology at the Hong Kong Institute of Education. Dr. Kong received a Doctor of Philosophy from the Department of Computer Science of the City University of Hong Kong. Dr. Kong publishes in the fields of information technology in education, information literacy education, collaboration in mobile learning and cognition, and technology in mathematics education. He is currently a member of the Steering Committee on the Strategic Development of Information Technology in Education that was set up by the Government of the Hong Kong SAR.

Panu Korpipää is a CEO of Finwe Ltd., formerly a senior research scientist at VTT. His research interests include enabling situation-aware and sensor-based interaction and applications in mobile computing. He received his doctoral Dr Tech in information processing from the University of Oulu.

Peter Kroeger has had a keen interest in computers and software development since 1985. He graduated from the University of Waterloo with an Honours Bachelor of Mathematics in Computer Science and a minor in Fine Arts (1993). His experience includes nine years at Corel Corporation where he was a project leader producing Web technology features in most of Corel's products. Since Corel, Peter has worked as a software architect and development manager. Currently Peter is a software architect at BRYTECH engaged in medical client-server and human interface technology. Peter founded the Software Professionals Society of Canada in 2003.

Fabrizio Lamberti received his degree in computer engineering and his PhD degree in software engineering from the Politecnico di Torino, Italy, in 2000 and 2005, respectively. He has published a number of technical papers in international journal and conferences in the areas of mobile and distributed computing, wireless networking, image processing and visualization. He has served as a reviewer and program or organization committee member for several conferences. He is member of the editorial advisory board of international journals. He is a member of the IEEE Computer Society.

James Landay is a professor in computer science & engineering at the University of Washington, specializing in HCI. Previously he was director of Intel Research Seattle, a lab exploring ubiquitous computing. His current interests include Automated Usability, Demonstrational Interfaces, Ubicomp, Design Tools, and Web Design. Landay received his BS in EECS from Berkeley in 1990 and his PhD in CS from Carnegie Mellon in 1996. His dissertation was the first to demonstrate sketching in UI design tools. He was also co-founder of NetRaker. In 1997 he joined the faculty in EECS at Berkeley, leaving as an associate professor in 2003.

Hyowon Lee is a postdoctoral researcher in the Centre for Digital Video Processing at Dublin City University where he obtained a PhD in user-interface design issues in keyframe-based video content browsing in 2001. Previously he obtained a BEng in computer science in Soong-Sil University, Korea in 1995, then a MSc in information & library studies at the Robert Gordon University, UK (1996). His area of research is the interaction design for multimedia retrieval systems such as image & video searching/browsing applications and photo & video blogging applications, and the evaluation and user studies involving these applications.

Jaakko Lehikoinen is a research team leader in the Nokia Research Center. He has been with Nokia since 1999 mostly working on and managing user-centered design oriented projects and programs. As a sociologist, he has been focusing on developing user research methodologies in such a way that the results can be utilized in concept creation and design. In addition to privacy, Jaakko's main research interests are social interaction and user experience design.

Rock Leung is a PhD student in the University of British Columbia's Department of Computer Science. He is pursuing research in human-computer interaction, focusing on interface design, mobile computing, and universal usability. His research interests include exploring how to improve computer interfaces and designing innovative assistive technologies for populations with special needs.

Dieter Leyk is professor at the German Sports University Cologne and the head of the Department IV–Military Ergonomics and Exercise Physiology at the Central Institute of the Ferderal Armed Forces Medical Services Koblenz. He studied both medicine and sports sciences and received the MD and PhD degrees. His main research topics are performance diagnostics, exercise physiology, human psychological performance and ergonomics.

Yang Li is a research associate of Computer Science & Engineering at the University of Washington. He was a postdoctoral researcher in EECS at UC Berkeley. He received a PhD in computer science from the Chinese Academy of Sciences in 2002. His research focus is primarily in user interface design tools, activity-based computing, and pen-based user interfaces.

Min Lin is a PhD candidate at UMBC. He received his MS in information systems from UMBC in 2002 and his MS in biology from the University of Science and Technology of China in 1995. Currently, he is a Usability Engineer at Hillcrest Laboratories Inc., where he is responsible for planning, conducting, and analyzing user studies for both hardware and software products. Before joining Hillcrest Labs, he had multiple years of research experience in both cognitive science and human-computer interaction.

Gitte Lindgaard, a full professor in psychology at Carleton University in Ottawa, holds the NSERC/Cognos chair in user-centred design in the Human Oriented Technology Lab (HOTLab) of which she is director. Formerly, as the principal scientist, head of the Human Factors Division at Telstra Research Laboratories, Australia before joining Carleton University, her job was to ensure that all Telstra's systems, services, and products were usable. Her research interests include human decision making, affective- as well as multimodal/multimedia computing, and graph comprehension. She has contributed several hundred peer-reviewed articles and book chapters. She serves on several HCI journal editorial boards.

Jukka Linjama, born in 1959, received his Dr Tech in electrical engineering in 1994, from Helsinki University of technology. His research interests span from acoustics, structural dynamics, and human perception of sound and vibration, to human-computer interaction, haptics and multi-modal interaction. Currently he acts as technology architect at Nokia Technology Platforms, responsible for specifying sensor technology and interaction solutions for future user interfaces in mobile devices.

Yanfang Liu leads the Strategic Design Research team within Motorola's Consumer Experience Design organization. Her team is responsible for design research on physical and digital aspects of mobile devices targeted to Asian markets, including Asian character entry. In her current role, Yanfang has contributed to several innovative character entry solutions that have been successfully delivered to market. Prior to joining the mobile devices business, Yanfang was a member of Motorola Labs' Human Interface Lab. Yanfang received her PhD in applied psychology from the Institute of Psychology at the Chinese Academy of Sciences in 1998.

Bob Longworth has recently completed his Honours degree in computer science at the University of New Brunswick. During his university career he completed five coop work terms three of which were with the National Research Council of Canada (NRC). Bob has interests in pharmaceutical drug design and has developed an interest toward Mobile HCI through his work at the NRC. Bob was also actively involved with his faculty in several ways. He was a peer mentor for three years, president of the Computer Science Association for two years, student representative on the faculty council for two years and was grad class rep.

Rosemarijn Looije received her masters degree in artificial intelligence/man machine interaction in 2006 from the University of Groningen. She is now a research member of the Intelligent Interface group at TNO Defence, Security and Safety in Soesterberg, The Netherlands. Her research interests lie within human-computer interaction, most notably the usability of devices, the dialogue with devices, and the influence of affect on the dialogue.

Saturnino Luz holds a PhD degree in informatics from the University of Edinburgh. Over the past years, he has worked on the development of human-computer interface technology in the areas of mobile systems and computer-supported collaborative work. His research interests also include natural language processing, information visualisation and retrieval, and machine learning. Dr Luz has served in the program committee of several international conferences and the editorial board of an international journal. He has been a member of the Association for Computing Machinery (ACM) since 1994 and contributes regularly to ACM's Computing Reviews.

Pattie Maes is an associate professor in MIT's Program in Media Arts and Sciences. She founded and directs the Media Lab's Ambient Intelligence research group, which focuses on linking the information world into the physical world around us. Previously, she founded and ran the Software Agents group. Prior to joining the Media Lab, Maes was a visiting professor and a research scientist at the MIT Artificial Intelligence Lab. She holds bachelor's and PhD degrees in computer science from the Vrije Universiteit Brussel in Belgium. Her areas of expertise are human-computer interaction and artificial intelligence. Maes is the editor of three books, and is an editorial board member and reviewer for numerous professional journals and conferences. She has received several awards: *Newsweek* magazine named her one of the "100 Americans to watch for" in the year 2000; TIME Digital selected her as a member of the Cyber-Elite, the top 50 technological pioneers of the high-tech world; the World Economic Forum honored her with the title "Global Leader for Tomorrow"; Ars Electronica awarded her the 1995 World Wide Web category prize; and in 2000 she was recognized with the "Lifetime Achievement Award" by the Massachusetts Interactive Media Council.

Massimiliano Margarone is a research consultant at the Department of Biophysical and Electronic Engineering of the University of Genova. He received his masters degree in electronic engineering from the University of Genova, in 1999, with a thesis on asynchronous digital filter development for wireless devices. His current research interests include Web usability, human-computer interaction, and mobile computing.

Masood Masoodian holds a PhD in computer science from the University of Waikato, New Zealand. His research interests include human-computer interaction, interaction design, visualisation, mobile computing and usability. Dr Masoodian has been involved in design, development, and evaluation of numerous graphical user interfaces and visualisation systems for devices ranging from large interactive displays to small handheld devices. He has served as the programme chair, programme committee member, and reviewer for many international conferences and scientific journals in HCI.

Mark Matthews is a PhD candidate with the Department of Computer Science at Trinity College Dublin. His research interests include the design of mobile software for therapeutic activities and developing new approaches to user-centred design for adolescent mental health. He is interested in the design of technology to engage adolescents between therapy sessions and in the use of technology, such as computer games, to develop therapeutic relationships.

Peter Michael Melliar-Smith is a professor in the Department of Electrical and Computer Engineering at the University of California, Santa Barbara. His research interests include distributed computing, network protocols, and fault tolerance. He has published more than 240 conference and journal papers.

Eduardo Mena earned a BS degree in computer science from the University of the Basque Country and a PhD degree in computer science from the University of Zaragoza. He is an associate professor in the Department of Computer Science and Systems Engineering at the University of Zaragoza, Spain. For a year he was a visiting researcher in the Large Scale Distributed Information Systems Laboratory at the University of Georgia. He leads the Distributed Information Systems research group at his university and is a member of the Interoperable Database group at the University of the Basque Country. His research interest areas include interoperable, heterogeneous and distributed information systems, semantic Web, and mobile computing. His main contribution is the OBSERVER project. He has published several papers in international journals, conferences and workshops, including a book on ontology-based query processing. He also has served as a referee for several international journals, and as a program committee member for several international conferences and workshops.

Kristijan Mihalic is a researcher in the HCI & Usability Unit at the University of Salzburg. He received a master's degree in the interdisciplinary field of communication sciences and computer science and holds a doctorate in HCI from the University of Salzburg, Austria. He was co-chair of the MobileHCI 2005 conference, and has co-organized several workshops on context and mobile HCI. Kristijan is involved in several national and international research and industry projects. In addition, he has a sound background in software development and several years of experience in teaching human-computer interaction and telecommunication technologies at national and international institutions. His primary research involves mobile and contextual interfaces, social aspects of mobile systems, and ambient intelligence. He lives on coffee.

John Miller has over 40 years experience in telecommunications and electronics design. He began his career with British Telecom and has held engineering and management positions with a number of companies including Canadian Marconi and Mitel. After completing his PhD in 1988 at Southampton University in England, he returned to Canada to lead a team at the University of Ottawa Heart Institute designing electronics for an artificial heart assist device. He received the 1993 Medforte Innovation Award, American Society of Artificial Organs, for his work on transcutaneous energy transfer. Currently John is VP of product development at BRYTECH Inc.

Nikola Mitrovic earned an MPhil degree in computer science from the University of Zaragoza, Spain. He is currently working toward a PhD degree in the area of intelligent and adaptive user interfaces. Nikola's research interest areas include interoperable, heterogeneous and distributed information systems and mobile computing. He has published several papers in international conferences and workshops and has served as a reviewer for several international conferences.

Michelle Montgomery Masters is a researcher in the Department of Computing and Information Sciences at the University of Strathclyde in Glasgow, Scotland, and is usability director for PyrusMalus™ software design company. She has expertise in designing and conducting usability evaluations on web, desktop and mobile applications. Michelle has a history of research into computer support for education, her current research focus is designing mobile learning systems. Michelle also acts as reviewer for many of the leading HCI publications and conferences. Prior to Strathclyde, Michelle was a researcher at The University of Glasgow.

Louise Elizabeth Moser is a professor in the Department of Electrical and Computer Engineering at the University of California, Santa Barbara. Her research interests span the fields of computer networks, distributed systems, and software engineering. She has published more than 225 conference and journal publications.

Wafaa Abou Moussa is a PhD student at the Research Institute in Computer Science of Toulouse–France (IRIT). Currently he conducts research on distributed 3D interaction and model-driven simulation of mobile mixed systems. He received his electronics engineering diploma from the "Lebanese University" in 2004 and his masters degree in computer science from the "Paul-Sabatier University." France in 2005.

Roderick Murray-Smith received degrees (BEng '90, PhD '94) from the University of Strathclyde, UK. He worked at Daimler-Benz research labs in Berlin from 1990-97, was a visiting researcher in the Dept. of Brain & Cognitive Sciences at MIT in 1994, a research fellow in the Dept. of Mathematical Modeling at the Technical University of Denmark from 1997-99, and since 1999 has been a reader at

Glasgow University in the Department of Computing Science. In 2001 he took up a joint (senior researcher) position at the Hamilton Institute, at NUI Maynooth, Ireland. He heads the Dynamics & Interaction research group, and his interests are in gesture recognition, mobile computing, manual control systems, Gaussian processes, and machine learning.

Bojan Musizza received his BSc in 2003 from the Faculty of Electrical Engineering, University of Ljubljana. Currently he is a PhD student working at the "Jozef Stefan" Institute at the Department of Systems and Control. His research interests are complex oscillatory systems and their interactions. His work mainly focuses on detecting interactions between cardiorespiratory oscillations and brain waves during anaesthesia in humans.

Mark Neerincx is head of the Intelligent Interface group at TNO Defence, Security and Safety, and professor in Man-Machine Interaction at the Delft University of Technology. He has extensive experience in applied and fundamental research. Important results are (1) a cognitive task load model for task allocation and adaptive interfaces, (2) models of human-machine partnership for attuning assistance to the individual user and momentary usage context, (3) cognitive engineering methods and tools, and (4) a diverse set of usability "best practices". He has been involved in the organisation of conferences, workshops and tutorials to discuss and disseminate human factors knowledge.

Shigueo Nomura is presently a JSPS postdoctoral fellow in the laboratory for Theory of Symbiotic Systems, Kyoto University, Japan. He received his BSc, BE, and MSc degrees in 1988, 1992, and 2002, respectively, from ITA at the Aerospace Technical Center, EPUSP at University of São Paulo, and FEELT at Federal University of Uberlândia in Brasil. He won a doctoral scholarship grant from the Japanese Government (Monbukagakusho) in 2002 and received his doctoral degree in 2006 from the Graduate School of Informatics, Kyoto University. His current research interests include human computer interaction, echolocation, pattern recognition, neurocomputing, and morphological image analysis.

Michael O'Grady is an SFI postdoctoral research fellow within the School of Computer Science & Informatics at University College Dublin (UCD). Prior to joining the school, he worked in the commercial software and telecommunications industries. After completing a postdoctoral fellowship in the area of pervasive computing funded by the Irish Research Council for Science and Engineering Technologies (IRCSET), he joined the Adaptive Information Cluster (AIC) group at UCD. His research interests include Ambient Intelligence, mobile multimedia and intelligent systems. He has published some 40 papers in international journals and conferences.

Gregory O'Hare is an SFI investigator and senior lecturer in the School of Computer Science & Informatics at University College Dublin. He was previously a member of faculty at the University of Manchester Institute of Science and Technology (UMIST), UK. He is director of the PRISM (Practice and Research in Intelligent Systems & Media) Laboratory within the School of Computer Science & Informatics. His research focuses upon multi-agent systems (MAS) and mobile & ubiquitous computing. He has published over 200 journal and conference papers in these areas together with two text books.

Kate Oakley completed her doctorate in psychology at Carleton University in 1998. Her dissertation, *Age and Individual Differences in the Realism of Confidence*, focused on decision making and led to a pilot research grant for new investigators from the Alzheimer's Association. Dr. Oakley was subsequently awarded an ORNEC two-year post-doctoral fellow position at the Human Oriented Technology Lab (HotLab), Carleton University, to develop computerized assessment tools in dementia. Currently

NSERC/Cognos research associate at the HOTLab, her research focuses on decision making, and graph comprehension for mobile technology. Kate is also part-time research scientist at the Élisabeth Bruyère Research Institute, Ottawa.

Marianna Obrist graduated in communication and political science at the University of Salzburg. She is a researcher in the HCI & Usability Unit of the ICT&S Center at the University of Salzburg. The focal point of her research lies in human-computer interfaces, user-centered design of interactive services, and in particular the user involvement into the development of new products/systems. She is involved in several research projects concerned with the study of the home environment and the analysis of user requirements for interactive TV. Since September 2006 she is working within the CITIZEN MEDIA project, focusing on co-creation and user experience evaluation in the new user driven media landscape. She was part of the organization team for the MobileHCI 2005 conference. She wrote her PhD on self-motivated adaptation and innovation of interfaces in the context home.

Gustav Öquist is a researcher at the Department of Clinical Neuroscience at Karolinska Institutet, Sweden. He is a computational linguist and his interests include novel concepts and solutions for information access with focus on language, usability, accessibility and mobility. He previously studied at Uppsala University and wrote his PhD thesis about evaluation of readability on mobile devices.

Jeni Paay is an assistant professor at Aalborg University, Denmark, in the Department of Computer Science, where she is a member of the Human-Computer Interaction (HCI) research group. She has completed a multi-discipline PhD in HCI and Architecture from The University of Melbourne, Australia. Her research interests include: pervasive computing in the built environment and the interplay between social interactions and architectural space; indexical interaction design for context-aware mobile computer systems; the representation of context in interface design; and HCI design methods.

Ioanna Papadimitriou was born in Athens, Greece (1978) and holds a degree in early childhood education from the University of Patras (2002), and an MSc in didactics of science (2006). Ioanna is currently a PhD candidate at the same University. Ioanna's research interests include the integration of computers in education and the educational uses of mobile technologies.

Hyungsung Park is a PhD candidate at the Korea National University of Education. His research interests are in using mobile devices for learning, designing educational games, and user interfaces in educational contexts. He has recently published two articles entitled *Development of learning contents in game to support a mobile learning* and *The analysis of the Knowledge Construction Types in Educational on-line Games on the Basis of the Levels of the Self-Regulated Learning* in Korea.

Sanghyun Park is a junior researcher in mobile UI R&D Lab at LG Electronics Inc. Before joining the LG Mobile UI lab in 2005, she also worked in UI Quality assurance Labs at Samsung Electronics Inc. She received an MS in industry engineering from the Korea Advanced Institute of Science and Technology (KAIST) in 2004 and a BS from SungKyunKwan University in 2003. Her research interests center on the user experience design and user research using ethnographic techniques with mobile contexts and industrial environments. She is also interested in physical user interface design and usability evaluation for UI designers to offer valuable user experiences.

Shwetak N. Patel is a PhD student in the College of Computing and GVU Center at the Georgia Institute of Technology. He is also an assistant director of the Aware Home Research Initiative. His research focuses on context-aware mobile phone applications and technology to support ubiquitous computing applications. He received his BS in computer science from the Georgia Institute of Technology. He's a member of the IEEE Computer Society and the ACM.

Antti Pirhonen holds a PhD with a major in educational sciences and minor in information systems. He is a senior researcher in the Department of Computer Science and Information Systems at the University of Jyväskylä Finland. Antti was a visiting researcher at the University of Glasgow's Department of Computing between 2000-2001 and again in 2005. He is the scientific leader of a number of projects funded by the Finnish Funding Agency for Technology and Innovation (Tekes) and the Finnish ICT industry. His research interests concern mobile applications, the use of non-speech audio in user-interfaces, multimodal interaction, and design methodologies.

Frank E. Pollick obtained a PhD in cognitive sciences in 1991 from UC Irvine. From 1991-97 he was a researcher at ATR in Kyoto, Japan. Since 1997 he has been in the Department of Psychology at the University of Glasgow. His research interests include the recognition of human movement.

Bhiksha Raj joined MERL as a staff scientist. He completed his PhD from Carnegie Mellon University (CMU) in May 2000. Dr. Raj works mainly on algorithmic aspects of speech recognition, with special emphasis on improving the robustness of speech recognition systems to environmental noise. Dr. Raj has over seventy five conference and journal publications and is currently writing a book on missing-feature methods for noise-robust speech recognition.

Tapani Rantakokko is a senior SW designer at Finwe Ltd., formerly a research scientist at VTT, and a doctoral student in the University of Oulu's Department of Electrical and Information Engineering. His research interests include mobile user interfaces and pervasive computing. He received his MSc in information processing from the University of Oulu.

Dimitrios Raptis was born in Chalkis, Greece (1980). He holds a degree in electrical and computer engineering from the University of Patras, Greece and is currently a PhD candidate at the same University pursuing doctoral research in the area of mobile usability. His research interests include methods for design of mobile applications, context aware computing, and usability of mobile devices.

Janet Read is a senior lecturer at the University of Central Lancashire in Preston, UK. She is well known for her work on child computer interaction and directs the ChiCI research group at her host institution. One of her areas of interest is in the use of mainstream evaluation methods with children and this research has motivated her interest in Wizard of Oz studies.

Michael Rohs is a senior research scientist with Deutsche Telekom Laboratories at TU Berlin. His primary research interest is in mobile and pervasive user interfaces, with a focus on the integration of physical and virtual aspects of the user's environment. He developed a visual marker system for camera phones and marker-based interaction techniques for passive media as well as electronic displays. He received a PhD in computer science from ETH Zurich, Switzerland, where he was a research assistant at the Institute for Pervasive Computing.

Anxo Cereijo Roibás is senior lecturer at the University of Brighton, visiting lecturer at Westminster University, at the Politecnico di Milano and the National Institute of Design (India). His expertise resides in the user experience in pervasive communication systems. He has been HCI manager at the

Mobile Internet Services Provider, HiuGO SpA, and User Experience Consultant for Vodafone and he has collaborated with the Nokia Research Center. He has coordinated ethnographic research addressing the future use of mobile phones as multimedia tools in collaboration with the Vodafone Group Foundation and the British Royal Academic of Engineering. He has been British Telecom Fellow at the BT IT Mobility Research Centre. Anxo is Events' Chair of the British HCI Group.

Jose A. Royo earned a BS degree in computer science from the University of Zaragoza, Spain, in 2001. In his dissertation he presented the advantages of mobile knowledge driven agents in distributed and heterogeneous systems. He is currently the system analyst of the Electronic Engineering and Communications Department at the University of Zaragoza, Spain. His research interests include: mobile agents; mobile computing; interoperable, heterogeneous and distributed information systems; and semantic Web. He has published several papers in international conferences and workshops, and also has served as a reviewer for some conferences, workshops and journals.

Andrea Sanna graduated in electronic engineering in 1993, and received a PhD degree in computer engineering in 1997, both from Politecnico di Torino, Italy. Currently he has an assistant professor position in the 2nd engineering faculty. He has authored and co-authored several papers in the areas of computer graphics, virtual reality, parallel and distributed computing, scientific visualization and computational geometry. Andrea Sanna is currently involved in several national and international projects concerning grid, peer-to-peer, and distributed technologies. He is a member of ACM and serves as reviewer for a number of international conferences and journals.

Giuseppe Santucci graduated in electrical engineering from the University of Rome "La Sapienza", Rome, Italy, on 1987. Since 1987 he teaches courses in computer science at Italian universities. From 1987 to 1991 he was research assistant at the University of Roma "La Sapienza", where he is now associate professor in the Department of Computer Science. His main research activity concerns both theoretical and practical aspects of visual query language and user interfaces. He is a member of the Steering Committee of the International Workshop on Advanced Visual Interfaces (AVI). He is a member of the Institute of Electrical and Electronics Engineers (IEEE).

Christopher M. Schlick received the Dipl.-Ing. degree in 1992 from Berlin University of Technology, Germany, the Dr.-Ing. and the Habilitation degrees from RWTH Aachen University, Germany, in 1999 and 2004, respectively. He is currently a full professor at the RWTH Aachen Faculty of Mechanical Engineering, where he is the director of the Institute of Industrial Engineering and Ergonomics. He received merits of honor from the German Human Factors Society GfA, RWTH (Borchers insignia 1999) and the Holste Foundation (Holste Price 2004). His research interests include the design and simulation of work and business systems, human-machine systems, and ergonomics.

Bent Schmidt-Nielsen has been developing and testing speech recognition and speech interfaces for automobile applications for the last dozen years. This includes seven years at Dragon Systems and six years at Mitsubishi Electric Research Laboratories. The primary thrust of his work is in the usability of speech interfaces for automobiles. Bent has worked building easy to use computer applications for the last 25 years. However, his academic training is in biology from the University of California at San Diego and the Massachusetts Institute of Technology.

Andrew Sears is a professor and chair of the Information Systems Department at UMBC. Dr. Sears' research explores issues related to human-centered computing with recent projects investigating issues associated with accessibility, mobile computing, and the difficulties IT users experience as a result of their work environments or tasks. He is co-editor of the Human-Computer Interaction Handbook and co-editor-in-chief of the ACM *Transactions on Accessible Computing.* He earned his BS in Computer Science from Rensselaer Polytechnic Institute and his PhD in computer science with an emphasis on human-computer interaction from the University of Maryland, College Park.

Sabine Seymour founded Moondial Inc., which focuses on "the next generation wearables" and the intertwining of aesthetics and function. Sabine introduced the course "Fashionable Technology" at Parsons in New York. She currently curates the exhibition "Cyborgs: Me or Machine" in England. Sabine is writing her PhD dissertation dealing with design and innovation in wearables. She received an MBA from the University of Economics in Vienna and Columbia University in New York and an MPS in Interactive Telecommunications from NYU'S Tisch School of the Arts.

John Sharry is a psychotherapist at the Department of Child and Family Psychiatry, at the Mater Hospital, Dublin, Director of the Parents Plus Charity and was formerly leader of the Therapeutic Technologies research group at Media Lab Europe in Dublin. His specific research interest is exploring the potential of computer technology in the field of psychotherapy with children and adolescents. John is author of nine books including three professional psychotherapy books, *Solution Focused Groupwork* (Sage, 2001), *Becoming a Solution Detective* (BT Press, 2001; Haworth 2003) and *Counselling children adolescents and Families* (Sage, 2004).

Jennifer G. Sheridan is director of BigDog Interactive, a company which creates bespoke interactive installations. Her interest is in digital live art, intersection of human-computer interaction and live art. She directs, designs and develops interactive installations and performance events using sensors and mobile and embedded physical computing technologies to mediate wittingness. She received an MS in HCI from the Georgia Institute of Technology, USA and a PhD in computer science from Lancaster University, UK.

Takayuki Shiose received his BE, ME and PhD degrees in 1996, 1998 and 2003, respectively, from the Faculty of Precision Engineering, Kyoto University. He was engaged as a research fellow with the Japan Society for the Promotion of Science from 1998 to 2000, and as a research associate with the Graduate School of Science and Technology, Kobe University from 2000 to 2002. Also, he was a visiting researcher at ATR Network Informatics Labs from 2002 to 2006. Now, he is an assistant professor at the Graduate School of Informatics, Kyoto University. His current research interests are in assistant systems for a proficient-skill transfer from the viewpoint of ecological psychology.

Katie A. Siek is an assistant professor at the University of Colorado at Boulder in the Department of Computer Science. Her primary research interests are in human computer interaction, health informatics, and ubiquitous computing. More specifically, she is interested in how technology affects medical interventions for underserved populations. Prior to her appointment at Colorado, she completed her PhD and MS at Indiana University, Bloomington in computer science and her undergraduate education in computer science at Eckerd College.

Alexander Sievert is a research associate at the Department of Physiology and Anatomy of the German Sport University Cologne in Cologne, Germany (DSHS). He obtained a Diplom-Sportlehrer (MS) degree and is currently working towards a doctorate in sports sciences. He began working for the DSHS in the area of human physical and psychological performance in 2004. He has also been working as a freelance consultant for local and international companies in the automotive industry and service providers for IT related Education, Training and Development. His main fields of research include work/exercise physiology, and ergonomics as well as human psychological performance.

Mikael B. Skov is associate professor at the Department of Computer Science at Aalborg University, Denmark. His professional interests include various aspects of human-computer interaction (HCI). He primarily conducts research within usability testing and engineering and within design and evaluation of mobile devices. He is involved in research projects on the design and evaluation of awareness-based mobile technologies and is also interested in social aspects of information technology design, evaluation, and use.

Alan Smeaton is a professor of computing and principal investigator in the Adaptive Information Centre at Dublin City University. His background is in text-based information retrieval and his research interests cover all aspects of the management of multimedia information.

Chiwon Song is a junior research fellow at the Mobile Handset R&D Center of LG electronics Company in Seoul, Korea. He received his BS in industrial engineering in 2003 from Ajou University, and MS in industrial engineering in 2003 and 2005, respectively, from Korea Advanced Institute of Science and Technology (KAIST). His research interests include the usability of physical user interfaces (PUI) on mobile devices and the features or areas to extend mobile device user's user experience.

Jan Stage is associate professor at the Department of Computer Science at Aalborg University, Denmark. He conducts research in usability testing and evaluation, design of human-computer interaction, design and evaluation of mobile systems, methods for analysis and design in system development and object-oriented analysis and design. He is currently involved in the following research activities: The USE Project (Usability Evaluation and Software Design: Bridging the Gap), the COST Project MAUSE (Towards the Maturation of Information Technology Usability Evaluation) and a high technology network on mobile systems.

Hanna Stelmaszewska is a PhD student and teaching assistant in the School of Computing Science at Middlesex University, where she is a member of the Interaction Design Centre. Her research interests include applying qualitative methods to investigate users' experience with interactive technology, user interaction and usability of interactive systems, which leads to gaining a better understanding of users' needs and support the design of systems. For her PhD she is evaluating users' experience with camera phones used for social interaction in co-present settings.

Adrian Stoica was born in Braila, Romania (1978). He obtained a degree in economic informatics from the University of Dunãrea de Jos of Galaţi, Romania, and is currently a PhD candidate at University of Patras, Greece, pursuing doctoral research in the area of ubiquitous computing. His research interests include the design of middleware for mobile context sensitive computing, interaction design for mobile devices, usability of mobile devices, and distributed user modeling in ubiquitous systems.

Jan Willem Streefkerk has a background in cognitive psychology. Currently he is a PhD student at TNO Defence, Security and Safety and Delft University of Technology in The Netherlands. His research interests focus on the design and evaluation of mobile, adaptive systems and their application in professional, critical domains. Previously, he worked as a lecturer on usability and research methods at The Hague University. He is an active member of CHI Nederland, the Dutch chapter of ACM SIGCHI.

Masanori Sugimoto received BEng, MEng, and Dr Eng degrees from the University of Tokyo, Japan, in 1990, 1992, and 1995, respectively. Currently, he is an associate professor at the Department of Frontier Informatics, Graduate School of Frontier Sciences, University of Tokyo. His research concern is related to human-computer interaction, especially mixed reality, mobile computing, computer supported collaborative work/learning, intelligent systems, and so on. Dr. Sugimoto is a member of ACM, IEEE, IEICE, JSAI, IPSJ, VRSJ, and JCSS.

Khai N. Truong is an assistant professor in the Department of Computer Science at the University of Toronto. His research interests include human-computer interaction and ubiquitous computing. Specifically, he focuses on the design and development of ubiquitous computing applications and then the evaluation of these systems' impact on daily life. He received his PhD in computer science from the Georgia Institute of Technology.

Nikolaos Tselios was born in Athens, Greece (1973). He obtained his Phd (2002) and a degree (1997) in electrical and computer engineering from the University of Patras, Greece. Currently, he is an adjunct lecturer and postgraduate fellow at the University of Patras, Greece. His main interests are related to design and evaluation of software interaction systems (with an emphasis in educational systems), design and evaluation of distance learning systems, cognitive models, context-aware computing systems and artificial intelligence techniques for human computer interaction.

Maria de Fátima Q Vieira Turnell received her Degree in physics from the UFPE in Brazil, in 1975, her masters degree in electrical engineering in 1979 from UFPB, and a PhD in electrical engineering from Bradford University, in the UK, in 1986. She is a senior lecturer at the Electrical Engineering Department of UFCG since 1977. She works in the undergraduate and graduate levels both in the Electrical Engineering Department and the Computer Science Department. Her main research interest is the design and evaluation of human interfaces for automated systems and in particular for mobile devices.

Myra van Esch-Bussemakers holds a masters degree in cognitive science and a PhD in multimodal user interfaces. Her research interests include usability, adaptive mobile user interfaces, multimodal interfaces and innovative user interfaces in complex, high-demand environments. Both as a scientist and as a project manager at TNO Defence, Security and Safety in The Netherlands, she is involved in projects that are more research-oriented as well as projects that are more applied, for instance in developing decision aids for medical patients on the Internet.

William Wong is professor of human-computer interaction, and head of the Interaction Design Centre, Middlesex University, London, UK. Before joining academia, Dr. Wong was an Air Defence Controller, and head of the Systems and Communications Operations Branch, HQ Republic of Singapore Air Force. He joined the University of Otago, New Zealand, in 1992, where he set up and directed the Multimedia Systems Research Laboratory, Department of Information Science, leaving the University as associate professor in 2003. He has conducted research into naturalistic decision making, situation awareness, and representation design in emergency ambulance control in New Zealand, Australia, and the UK; in air traffic control in Europe, exploring 3-D and 4-D representation design, the representation of shared weather information; minimisation of change blindness using visual depth design techniques; and the design of control interfaces for hydro-electricity generation in Australia. Dr Wong gained his PhD in human-computer interaction from the University of Otago, NZ in 1999.

Keiji Yamanaka received his PhD degree in electrical and computer engineering from Nagoya Institute of Technology, Japan, in 1999. He is currently working in the Faculty of Electrical Engineering, Federal University of Uberlândia, Brazil, teaching undergraduate and graduate courses. He has several master and doctoral students under his supervision and his research interests include neuro-computing, artificial intelligence, and pattern recognition.

Nikoletta Yiannoutsou was born in Athens, Greece (1973), and received a degree in psychology (1994) from the University of Athens (NKUA), an MSc in intelligent tutoring systems (1996) from University of Nottingham, and a Phd in educational technology (2005) from the University of Athens (NKUA). Currently, she is an adjunct Lecturer at the University of Peloponnese. Her research interests focus on the role of new representational media in the learning process and on the design of innovative educational activities based on ICT.

Dong Yu joined Microsoft in 1998 and the Microsoft Speech Research Group in 2002. He holds a PhD in computer science from the University of Idaho, an MS degree in computer science from Indiana University / Bloomington, an MS degree in electrical engineering from the Chinese Academy of Sciences, and a BS degree (with honors) in electrical engineering from Zhejiang University (China). His research interests include speech processing, pattern recognition, and computer and network security. He has published dozens of papers in these areas. He is a senior member of IEEE.

Ioannis D. Zaharakis received a BSc in mathematics in 1992 and a PhD in software engineering in 1999, both from the University of Patras, Hellas. He is a member of the Hellenic AI Society and the Hellenic Mathematics Society. During 1993-1999, he was a researcher in the Educational Software Development Lab, University of Patras, where he was involved in the specification, design and implementation of intelligent tutoring systems. He has participated in several EU-funded and national R&D projects. He has authored one textbook for the Hellenic Open University and he is currently teaching in the Technological Educational Institute of Patras. In 2002, he joined CTI/RU3 as a Researcher / R&D Engineer, focusing on the formal specification of component-based ubiquitous computer systems.

Martina Ziefle is a psychologist and professor at RWTH Aachen University, Germany. Her research interests cover a broad field of key issues within human computer interaction, cognitive ergonomics and interface design. A main research topic is concerned with the user-centered interface design of small screen devices and the impact of user diversity for technical devices.

Index

D

E

F

G

H

N

O

P

Q

R

V

W

Z