

Commenced Publication in 1973

Founding and Former Series Editors:

Gerhard Goos, Juris Hartmanis, and Jan van Leeuwen

Editorial Board

David Hutchison

Lancaster University, UK

Takeo Kanade

Carnegie Mellon University, Pittsburgh, PA, USA

Josef Kittler

University of Surrey, Guildford, UK

Jon M. Kleinberg

Cornell University, Ithaca, NY, USA

Alfred Kobsa

University of California, Irvine, CA, USA

Friedemann Mattern

ETH Zurich, Switzerland

John C. Mitchell

Stanford University, CA, USA

Moni Naor

Weizmann Institute of Science, Rehovot, Israel

Oscar Nierstrasz

University of Bern, Switzerland

C. Pandu Rangan

Indian Institute of Technology, Madras, India

Bernhard Steffen

TU Dortmund University, Germany

Madhu Sudan

Microsoft Research, Cambridge, MA, USA

Demetri Terzopoulos

University of California, Los Angeles, CA, USA

Doug Tygar

University of California, Berkeley, CA, USA

Gerhard Weikum

Max-Planck Institute of Computer Science, Saarbruecken, Germany

Klaus Miesenberger Joachim Klaus
Wolfgang Zagler Arthur Karshmer (Eds.)

Computers Helping People with Special Needs

12th International Conference, ICCHP 2010
Vienna, Austria, July 14-16, 2010
Proceedings, Part II

Volume Editors

Klaus Miesenberger
University of Linz, Institute "Integriert Studieren"
Altenbergerstraße 49, 4040, Linz, Austria
E-mail: klaus.miesenberger@jku.at

Joachim Klaus
University of Karlsruhe (TH), Study Center for the Visually Impaired Students
Engesserstr. 4, 76131 Karlsruhe, Germany
E-mail: klaus@szs.uni-karlsruhe.de

Wolfgang Zagler
Vienna University of Technology, Institute "Integriert Studieren"
Favoritenstr. 11/029, A 1040, Vienna, Austria
E-mail: zw@fortec.tuwien.ac.at

Arthur Karshmer
University of San Francisco
2130 Fulton St., San Francisco, CA 94117, USA
E-mail: arthur@lakeland.usf.edu

Library of Congress Control Number: 2010929485

CR Subject Classification (1998): H.5.2, H.5, H.4, H.3, K.4, K.4.2

LNCS Sublibrary: SL 3 – Information Systems and Application, incl. Internet/Web and HCI

ISSN 0302-9743
ISBN-10 3-642-14099-8 Springer Berlin Heidelberg New York
ISBN-13 978-3-642-14099-0 Springer Berlin Heidelberg New York

This work is subject to copyright. All rights are reserved, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, re-use of illustrations, recitation, broadcasting, reproduction on microfilms or in any other way, and storage in data banks. Duplication of this publication or parts thereof is permitted only under the provisions of the German Copyright Law of September 9, 1965, in its current version, and permission for use must always be obtained from Springer. Violations are liable to prosecution under the German Copyright Law.

springer.com

© Springer-Verlag Berlin Heidelberg 2010
Printed in Germany

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India
Printed on acid-free paper 06/3180

Preface

Welcome to the Proceedings of ICCHP 2010!



We were proud to welcome participants from more than 40 countries from all over the world to this year's ICCHP.

Since the late 1980s, it has been ICCHP's mission to support and reflect development in the field of "Assistive Technologies," eAccessibility and eInclusion. With a focus on scientific quality, ICCHP has become an important reference in our field. The 2010

conference and this collection of papers once again fulfilled this mission.

The International Programme Committee, comprising 106 experts from all over the world, selected 147 full and 44 short papers out of 328 abstracts submitted to ICCHP. This acceptance ratio of about half of the submissions demonstrates our strict pursuit of scientific quality both of the programme and in particular of the proceedings in your hands.

An impressive number of experts agreed to organize "Special Thematic Sessions" (STS) for ICCHP 2010. These STS help to bring the meeting into sharper focus in several key areas. In turn, this deeper level of focus helps to collate a state of the art and mainstream technical, social, cultural and political developments.

Our keynote speakers highlighted issues of outstanding importance, such as "*Usable Web Accessibility*," areas which can be seen as breakthroughs in our field, like the development of 2D Braille displays *allowing "Maintaining an Overview Through Your Fingers*," and new and upcoming issues of interest for our field, like "*Weak and Silent Speech*." Additional keynote addresses included presentations from the *European Commission, UNESCO and ISO*.

With an interesting workshop programme, meetings, an exhibition including presentations and demonstrations of major software and assistive technology producers and vendors, ICCHP once again provided the framework for an international meeting place and a center of advanced information exchange.

ICCHP 2010 was held under the auspices of Heinz Fischer, President of the Federal Republic of Austria. We would also like to thank all supporters and sponsors.

Our special thanks go to those contributing toward putting this conference in place.

We thank the Austrian Computer Society for announcing and sponsoring the Roland Wagner Award on Computers Helping People with Special Needs.

The Austrian Computer Society decided in September 2001 to endow this award in honor of Roland Wagner, the founder of ICCHP.

The Roland Wagner Award is a biannual award in the range of €3000. It will be handed over on the occasion of ICCHP conferences.

Award Winners:

- Award 0: Roland Wagner on the occasion of his 50th birthday, 2001
- Award 1: WAI-W3C, ICCHP 2002 in Linz
- Special Award 2003: A Min Tjoa, Vienna University of Technology on the occasion of his 50th birthday
- Award 2: Paul Blenkhorn, University of Manchester, ICCHP 2004 in Paris
- Award 3: Larry Scadden, National Science Foundation, ICCHP 2006 in Linz
- Special Award 2006: Roland Traunmüller, University of Linz
- Award 4: George Kersher, Daisy Consortium

Once again we thank all those who helped in putting ICCHP in place and thereby supporting the AT field and a better quality of life for people with disabilities. Special thanks go to all our supporter and sponsors, displayed at:
<http://www.icchp.org/sponsors>.

July 2010

Klaus Miesenberger
Joachim Klaus
Arthur Karshmer
Wolfgang Zagler

Organization

General ICCHP 2010 Chair, H.J. Murphy, Founder, Former Director and Member Advisory Board of the Centre on Disabilities, California State University, Northridge, USA.

Programme Chairs

Karshmer, A.	University of San Francisco, USA
Burger, D.	INSERM, France
Suzuki, M.	Kyushu University, Japan
Tjoa, A. M.	Vienna University of Technology, Austria
Wagner, R.	University of Linz, Austria

Publishing Chairs

Klaus, J.	Karlsruhe Institute of Technology (KIT), Germany
Miesenberger, K.	University of Linz, Austria
Zagler, W.	Vienna University of Technology, Austria

Young Researchers Consortium Chairs

Alm, N.	Dundee University, UK
Archambault, D.	Université Pierre et Marie Curie, France
Fels, D.	Ryerson University, Canada
Fitzpatrick, D.	Dublin City University, Ireland
Kobayashi, M.	Tsukuba College of Technology, Japan
Morandell, M.	AIT Austrian Institute of Technology GmbH, Austria
Pontelli, E.	New Mexico State University, USA
Sinclair, R.	Microsoft, USA
Weber, G.	Technische Universität Dresden, Germany

Programme Committee

Abascal, J.	Euskal Herriko Unibertsitatea, Spain
Abou-Zahra, S.	W3C Web Accessibility Initiative (WAI), Austria
Abu-Ali, A.	Philadelphia University, Jordan
Andrich, R.	Polo Tecnologico Fondazione Don Carlo Gnocchi Onlus, Italy
Arató, A.	KFKI-RMKI, Hungary
Arató, P.	TU Budapest, Hungary
Ariwa, E.	London Metropolitan University, UK
Azevedo, L.	Instituto Superior Tecnico, Portugal

Batusic, M.	University of Linz, Austria
Baumann, K.	FH Johaneum, Austria
Bernareggi, C.	Università degli Studi di Milano, Italy
Bothe, H.-H.	Technical University of Denmark, Denmark
Bühler, C.	TU Dortmund University, FTB, Germany
Craddock, G.	Centre for Excellence in Universal Design, Ireland
Craven, J.	CERLIM, Manchester Metropolitan University, UK
Crombie, D.	European Adaptive Content Network (EUAIN), The Netherlands
Cummins Prager, M.	California State University Northridge, USA
Darvishy, A.	Züricher Hochschule für Angewandte Wissenschaften, Switzerland
Darzentas, J.	University of the Aegean, Greece
DeRuyter, F.	Duke University Medical Center, USA
Diaz del Campo, R.	Antarq Tecnosoluciones, Mexico
Dotter, F.	Universität Klagenfurt, Austria
Edwards, A.D.N.	University of York, UK
Emiliani, P.L.	Institute of Applied Physics "Nello Carrara", Italy
Engelen, J.	Katholieke Universiteit Leuven, Belgium
Evreinov, G.	University of Tampere, Finland
Gardner, J.	Oregon State University, USA
Gupta, G.	University of Texas at Dallas, USA
Hanson, V.	University of Dundee, UK
Harper, S.	University of Manchester, UK
Holzinger, A.	Graz University Hospital, Austria
Inoue, T.	National Rehabilitation Center for Persons with Disabilities, Japan
Ioannidis, G	IN2 search interfaces development Ltd, Germany
Jemni, M.	University of Tunis, Tunisia
Kalinnikova, L.	Pomor State University, Russia
Knops, H.	NG4All, The Netherlands
Koronios, A.	University of South Australia, Australia
Kouroupetroglou, G.	University of Athens, Greece
Kremser, W.	OCG, HSM, Austria
Lauruska, V.	Siauliai University, Lithuania
Leahy, D.	Trinity College Dublin, Ireland
Lunn, D.	University of Manchester, UK
Magnussen, M.	Stockholm University, Sweden
Mathiassen, N-E.	Danish Centre for Assistive Technology, Denmark
McKenzie, N.	DEDICON, The Netherlands
McMullin, B.	Dublin City University, Ireland
Mendelova, E.	Comenius University of Bratislava, Slovak Republic
Neveryd, H.	Lund University, Sweden
Normie, L.	GeronTech - The Israeli Center for Assistive Technology & Aging, Israel
Nussbaum, G.	KI-I, Austria
Ohnabe, H.	Niigata University of Health & Welfare, Japan

Ono, T.	Tsukuba University of Technology, Japan
Otogawa, T.	National Rehabilitation Center for Persons with Disabilities, Japan
Paciello, M.	The Paciello Group, USA
Panek, P.	Vienna University of Technology, Austria
Petrie, H.	University of York, UK
Petz, A.	University of Linz, Austria
Prazak, B.	Austrian Research Centres, Austria
Pühretmair, F.	KI-I, Austria
Quirchmayr, G.	University of Vienna, Austria
Raisamo, R.	University of Tampere, Finland
Rauhala, M.	Vienna University of Technology, Austria
Salminen, A.	KELA, Finland
Scadden, L.	Independent Consultant on Assistive Technology, USA
Schultz, T.	Universität Karlsruhe (TH), Germany
Sik Lányi, C.	University of Pannonia, Hungary
Snarud, M.	University of Agder, Norway
Soede, T.	Zuyd University, The Netherlands
Steiner, D.	University of Linz, Austria
Stephanidis, C.	University of Crete, FORTH-ICS, Greece
Stoeger, B.	University of Linz, Austria
Strauss, C.	University of Vienna, Austria
Sueda, O.	University of Tokushima, Japan
Svensson, H.	National Agency for Special Needs Education and Schools, Sweden
Takahashi, Y.	Toyo University, Japan
Tauber, M.	University of Paderborn, Germany
Traunmüller, R.	University of Linz, Austria
Trehin, P.	World Autism Organisation, Belgium
Treviranus, J.	University of Toronto, Canada
Vaspöri, T.	KFKI-RMKI, Hungary
Velasco, C.A.	Fraunhofer Institute for Applied Information Technology, Germany
Vigouroux, N.	IRIT Toulouse, France
Wagner, Gerold	Upper Austria University of Applied Sciences, Austria
Weber, H.	ITA, University of Kaiserslautern, Germany
Wöß, W.	University of Linz, Austria

Organizing Committee (University of Linz, Vienna Technical University)

Bernkopf, G.	Ossmann, R.	Keim, C.
Feichtenschlager, P.	Petz, A.	Neustätter, G.
Miesenberger, K.	Schult, C.	Vomastek, C.
Morandell, M.	Fuhrmann – Ehn, M.	Zagler, W.

Table of Contents – Part II

People with Specific Learning and Cognitive Problems: ICT, AT and HCI

Methodological Considerations for Involving SpLD Practitioners and Specialists in Designing Interactive Learning Systems	1
<i>Latifa Al-Abdulkarim, Areej Al-Wabil, Maha Al-Yahya, Abeer Al-Humaimedy, and Sara Al-Khudair</i>	
Involving Users in the Design of ICT Aimed to Improve Education, Work, and Leisure for Users with Intellectual Disabilities	5
<i>Emanuela Mazzone, Emmanuelle Gutiérrez y Restrepo, Carmen Barrera, Cecile Finat, Olga C. Santos, Jesús Boticario, Javier Moranchel, Jose Ramón Roldán, and Roberto Casas</i>	
PDA Software Aimed at Improving Workplace Adaptation for People with Cognitive Disabilities	13
<i>Alberto Ferreras, Juan-Manuel Belda, Ricard Barberà, Rakel Poveda, Miguel Urra, Nuria García, Miguel Tito, and Marta Valero</i>	
EasyICT: A Framework for Measuring ICT-Skills of People with Cognitive Disabilities	21
<i>Jan Dekelver, Tim Vannuffelen, and Joan De Boeck</i>	
Towards an Interactive Screening Program for Developmental Dyslexia: Eye Movement Analysis in Reading Arabic Texts	25
<i>Areej Al-Wabil and Maha Al-Sheaha</i>	
Developing a Multimedia Environment to Aid in Vocalization for People on the Autism Spectrum: A User-Centered Design Approach	33
<i>Areej Al-Wabil, Hadeel Al-Shabanat, Rawan Al-Sarrani, and Manal Al-Khonin</i>	
The Performance of Mouse Proficiency for Adolescents with Intellectual Disabilities	37
<i>Ting-Fang Wu, Ming-Chung Chen, and Chi-Fen Wu</i>	
When Words Fall Short: Helping People with Aphasia to Express	45
<i>Tom Koppenol, Abdullah Al Mahmud, and Jean-Bernard Martens</i>	

People with Motor Disabilities: HCI and Rehabilitation

MarkerMouse: Mouse Cursor Control Using a Head-Mounted Marker . . . <i>Rados Javanovic and Ian Scott MacKenzie</i>	49
Using Triaxial Acceleration Sensors as Input Devices for Disabled Persons <i>Matthias Söllner, Philipp Hofmann, and Josef Pösl</i>	57
Canceling Interference Caused by Tremors in Joystick Controller: Study Case in a Power Wheelchair <i>Ludmila Correa de Alkmin Silva, Fernanda Cristina Corrêa, Geraldo Gonçalves Delgado Neto, Vivianne V. Delgado, and Franco Giuseppe Dedini</i>	61
Developing Rehabilitation Robots for the Brain Injured <i>Paul Gnanayutham and Jennifer George</i>	69
A Rehabilitation Method with Visual and Haptic Guidance for Children with Upper Extremity Disability <i>Kup-Sze Choi, Chum-Ming Chow, and King-Hung Lo</i>	77
SSVEP Based Brain-Computer Interface for Robot Control <i>Rupert Ortner, Christoph Guger, Robert Prüeckl, Engelbert Grünbacher, and Günter Edlinger</i>	85

People with Motor Disabilities: How to Improve Text Input

Text Composing Software for Disabled People, Using Blink and Motion Detection <i>Philipp Hofmann, Matthias Söllner, and Josef Pösl</i>	91
Augmented and Alternative Communication System Based on Dasher Application and an Accelerometer <i>Isabel Gómez, Pablo Anaya, Rafael Cabrera, Alberto J. Molina, Octavio Rivera, and Manuel Merino</i>	98
Evaluation of WordTree System with Motor Disabled Users <i>Georges Badr and Mathieu Raynal</i>	104
Evaluation of SpreadKey System with Motor Impaired Users <i>Bruno Merlin and Mathieu Raynal</i>	112
An Integrated Text Entry System Design for Severe Physical Disabilities <i>Yun-Lung Lin, Ming-Chung Chen, Yao-Ming Yeh, and Chih-Ching Yeh</i>	120

Qanti: A Software Tool for Quick Ambiguous Non-standard Text Input	128
<i>Torsten Felzer, Ian Scott MacKenzie, Philipp Beckerle, and Stephan Rinderknecht</i>	
A Prototype Scanning System with an Ambiguous Keyboard and a Predictive Disambiguation Algorithm	136
<i>Julio Miró-Borrás, Pablo Bernabeu-Soler, Raul Llinares, and Jorge Igual</i>	
An Ambiguous Keyboard Based on “Character Graphical Association” for the Severely Physically Handicapped	140
<i>Julio Miró-Borrás, Pablo Bernabeu-Soler, Raul Llinares, and Jorge Igual</i>	
Deaf and Hard of Hearing People: Accessible Information and Education	
The Deaf and Online Comprehension Texts, How Can Technology Help?	144
<i>Simona Ottaviano, Gianluca Merlo, Antonella Chifari, Giuseppe Chiazzese, Luciano Seta, Mario Allegra, and Valentina Samperi</i>	
Extraction of Displayed Objects Corresponding to Demonstrative Words for Use in Remote Transcription	152
<i>Yoshinori Takeuchi, Hajime Ohta, Noboru Ohnishi, Daisuke Wakatsuki, and Hiroki Minagawa</i>	
E-Scribe: Ubiquitous Real-Time Speech Transcription for the Hearing-Impaired	160
<i>Zdenek Bumbalek, Jan Zelenka, and Lukas Kencl</i>	
SportSign: A Service to Make Sports News Accessible to Deaf Persons in Sign Languages	169
<i>Achraf Othman, Oussama El Ghou, and Mohamed Jemni</i>	
Synote: Accessible and Assistive Technology Enhancing Learning for All Students	177
<i>Mike Wald</i>	
Teaching English to Deaf Adults: “SignOnOne” – An Online Course for Beginners	185
<i>Marlene Hilzensauer</i>	

Deaf People: AT for Sign Language

A Web Based Platform for Sign Language Corpus Creation	193
<i>Davide Barberis, Nicola Garazzino, Elio Piccolo, Paolo Prinetto, and Gabriele Tiotto</i>	
Context Analysis of Universal Communication through Local Sign Languages Applying Multivariate Analysis	200
<i>Naotsune Hosono, Hiromitsu Inoue, and Yuji Nagashima</i>	
Toward Automatic Sign Language Recognition from Web3D Based Scenes	205
<i>Kabil Jaballah and Mohamed Jemni</i>	
Introducing Arabic Sign Language for Mobile Phones	213
<i>Hend S. Al-Khalifa</i>	
Sign Language Interpreter Module: Accessible Video Retrieval with Subtitles	221
<i>Primož Kosec, Matjaž Debevc, and Andreas Holzinger</i>	

Blind and Partially Sighted People: Mobility and Interaction without Sight

Real-Time Walk Light Detection with a Mobile Phone	229
<i>Volodymyr Ivanchenko, James Coughlan, and Huiying Shen</i>	
An Ultrasonic Blind Guidance System for Street Crossings	235
<i>Satoshi Hashino and Sho Yamada</i>	
Pedestrian Navigation System Using Tone Gradient and Robotic GIS	239
<i>Takafumi Ienaga, Yukinobu Sugimura, Yoshihiko Kimuro, and Chikamune Wada</i>	
MOST-NNG, an Accessible GPS Navigation Application Integrated into the MOBILE Slate Talker (MOST) for the Blind	247
<i>Norbert Márkus, András Arató, Zoltán Juhász, Gábor Bognár, and László Késmárki</i>	
Improving Computer Vision-Based Indoor Wayfinding for Blind Persons with Context Information	255
<i>YingLi Tian, Chucai Yi, and Aries Arditi</i>	
Computer Vision-Based Door Detection for Accessibility of Unfamiliar Environments to Blind Persons	263
<i>Yingli Tian, Xiaodong Yang, and Aries Arditi</i>	
Development and Installation of Programmable Light-Emitting Braille Blocks	271
<i>Makoto Kobayashi and Hiroshi Katoh</i>	

FLIPPS for the Blind - An Object Detector for Medium Distances Using Concerted Fingertip Stimulation	275
<i>Hans-Heinrich Bothe and Sermed Al-Hamdani</i>	
Camera Based Target Acquisition Augmented with Phosphene Sensations	282
<i>Tatiana G. Evreinova, Grigori Evreinov, and Roope Raisamo</i>	
A Mobile Phone Application Enabling Visually Impaired Users to Find and Read Product Barcodes	290
<i>Ender Tekin and James M. Coughlan</i>	
A Model to Develop Videogames for Orientation and Mobility	296
<i>Jaime Sánchez, Luis Guerrero, Mauricio Sáenz, and Héctor Flores</i>	
Underestimated Cerebral Visual Impairment Requires Innovative Thinking When Using AT	304
<i>Michael Cyrus and Frank Lunde</i>	
Blind and Partially Sighted People: Rehabilitation and Social Skills	
Eye Tracking for Low Vision Aids - Toward Guiding of Gaze	308
<i>Yasuyuki Murai, Masaji Kawahara, Hisayuki Tatsumi, Iwao Sekita, and Masahiro Miyakawa</i>	
Assistive Technologies as Effective Mediators in Interpersonal Social Interactions for Persons with Visual Disability	316
<i>Sreekar Krishna and Sethuraman Panchanathan</i>	
Clothes Matching for Blind and Color Blind People	324
<i>Yingli Tian and Shuai Yuan</i>	
A Basic Inspection of Wall-Climbing Support System for the Visually Challenged	332
<i>Makoto Kobayashi</i>	
Makeup Support System for Visually Impaired Persons: Overview of System Functions	338
<i>Akihiko Hanafusa, Shuri Terada, Yuuri Miki, Chiharu Sasagawa, Tomozumi Ikeda, and Teruhiko Fuwa</i>	
Blind and Partially Sighted People: HCI	
WebTrax: Visualizing Non-Visual Web Interactions	346
<i>Jeffrey P. Bigham and Kyle Murray</i>	

A Context-Based Grammar Generation in Mixed Initiative Dialogue System for Visually Impaired	354
<i>Jaromír Plhák</i>	
Design and Development of Spoken Dialog Systems Incorporating Speech Synthesis of Viennese Varieties	361
<i>Michael Pucher, Friedrich Neubarth, and Dietmar Schabus</i>	
Methods for the Visually Impaired to Retrieve Audio Information Efficiently	367
<i>Atsushi Imai, Nobumasa Seiyama, Tohru Takagi, and Tohru Ifukube</i>	
Binocular Vision Impairments Therapy Supported by Contactless Eye-Gaze Tracking System	373
<i>Lukasz Kosikowski and Andrzej Czyżewski</i>	
Realization of Direct Manipulation for Visually Impaired on Touch Panel Interface	377
<i>Tatsumi Nishizawa, Daisuke Nagasaka, Hiroshi Kodama, Masami Hashimoto, Kazunori Itoh, and Seiji Miyaoka</i>	
Inspired by Sharp Vision: The Full HD Desktop Video Magnifier “Eagle”	385
<i>Maria Schiewe, Thorsten Völkel, and Dirk Kochanek</i>	

Blind and Partially Sighted People: Access to Mathematics

Non-sequential Mathematical Notations in the LAMBDA System	389
<i>Cristian Bernareggi</i>	
MathML to ASCII-Braille and Hierarchical Tree Converter	396
<i>Silvia Fajardo-Flores, Maria Andrade-Arechiga, Alfonso Flores-Barriga, and Juan Lazaro-Flores</i>	
Tactile Graphic Tool for Portable Digital Pad	403
<i>Thamburaj Robinson and Atulya K. Nagar</i>	
Spoken Mathematics Using Prosody, Earcons and Spearcons	407
<i>Enda Bates and Dónal Fitzpatrick</i>	
On Necessity of a New Method to Read Out Math Contents Properly in DAISY	415
<i>Katsuhito Yamaguchi and Masakazu Suzuki</i>	
Innovative Methods for Accessing Mathematics by Visually Impaired Users	423
<i>Giuseppe Nicotra, Giovanni Bertoni, Enrico Bortolazzi, and Luciana Formenti</i>	

Blind and Partially Sighted People: Designing Haptic Interaction for a Collaborative World

ICCHP Keynote: Designing Haptic Interaction for a Collaborative World	431
<i>Gerhard Weber</i>	
Improving the Accessibility of ASCII Graphics for the Blind Students: Producing My Own Graphics	439
<i>Karin Müller and Angela Constantinescu</i>	
Development of a Musical Score System Using Body-Braille	443
<i>Satoshi Ohtsuka, Nobuyuki Sasaki, Sadao Hasegawa, and Tetsumi Harakawa</i>	
Analysis of Awareness while Touching the Tactile Figures Using Near-Infrared Spectroscopy	447
<i>Takayuki Shiose, Yasuhiro Kagiya, Kentaro Toda, Hiroshi Kawakami, Kiyohide Ito, Akira Takatsuki, and Akitoshi Seiyama</i>	
Development of Directly Manipulable Tactile Graphic System with Audio Support Function	451
<i>Shigenobu Shimada, Haruka Murase, Suguru Yamamoto, Yusuke Uchida, Makoto Shimojo, and Yutaka Shimizu</i>	
AutOMathic Blocks: Supporting Learning Games for Young Blind Students	459
<i>Arthur I. Karshmer and Ken Paap</i>	
Audio-Haptic Browser for a Geographical Information System	466
<i>Limin Zeng and Gerhard Weber</i>	
A Graphical Tactile Screen-Explorer	474
<i>Martin Spindler, Michael Kraus, and Gerhard Weber</i>	
Reading Braille and Tactile Ink-Print on a Planar Tactile Display	482
<i>Denise Prescher, Oliver Nadig, and Gerhard Weber</i>	
Enhancing Single Touch Gesture Classifiers to Multitouch Support	490
<i>Michael Schmidt and Gerhard Weber</i>	
User-Interface Filter for Two-Dimensional Haptic Interaction	498
<i>Wiebke Köhlmann, Francis Zinke, Maria Schiewe, and Helmut Jürgensen</i>	

Improving Screen Magnification Using the HyperBraille Multiview
 Windowing Technique 506
*Christiane Taras, Michael Raschke, Thomas Schlegel, Thomas Ertl,
 Denise Prescher, and Gerhard Weber*

**Blind and Partially Sighted People:
 Three-Dimensional Tactile Models for Blind
 People and Recognition of 3D Objects by Touch**

Three-Dimensional Tactile Models for Blind People and Recognition of
 3D Objects by Touch: Introduction to the Special Thematic Session . . . 513
Yoshinori Teshima

Models of Mathematically Defined Curved Surfaces for Tactile
 Learning 515
*Yoshinori Teshima, Tohru Ogawa, Mamoru Fujiyoshi,
 Yuji Ikegami, Takeshi Kaneko, Susumu Oouchi,
 Yasunari Watanabe, and Kenji Yamazawa*

Enlarged Skeleton Models of Plankton for Tactile Teaching 523
*Yoshinori Teshima, Atsushi Matsuoka, Mamoru Fujiyoshi,
 Yuji Ikegami, Takeshi Kaneko, Susumu Oouchi,
 Yasunari Watanabe, and Kenji Yamazawa*

Reproduction of Tactile Paintings for Visual Impairments Utilized
 Three-Dimensional Modeling System and the Effect of Difference in the
 Painting Size on Tactile Perception 527
Susumu Oouchi, Kenji Yamazawa, and Lorreta Secchi

Tactile Map Automated Creation System to Enhance the Mobility
 of Blind Persons-Its Design Concept and Evaluation through
 Experiment 534
*Kazunori Minatani, Tetsuya Watanabe, Toshimitsu Yamaguchi,
 Ken Watanabe, Joji Akiyama, Manabi Miyagi, and Susumu Oouchi*

Development of Tactile Graphics Production Software for
 Three-Dimensional Projections 541
*Mamoru Fujiyoshi, Takeshi Kaneko, Akio Fujiyoshi,
 Susumu Oouchi, Kenji Yamazawa, Yuji Ikegami,
 Yasunari Watanabe, and Yoshinori Teshima*

Comprehending and Making Drawings of 3D Objects by Visually
 Impaired People: Research on Drawings of Geometric Shapes by
 Various Methods of Projection 548
*Takeshi Kaneko, Mamoru Fujiyoshi, Susumu Oouchi,
 Yoshinori Teshima, Yuji Ikegami, Yasunari Watanabe, and
 Kenji Yamazawa*

Ageing: HCI Usability (HCI4AGING)

Human-Computer Interaction and Usability Engineering for Elderly (HCI4AGING):

Introduction to the Special Thematic Session	556
<i>Andreas Holzinger, Martina Ziefle, and Carsten Röcker</i>	
A Device to Evaluate Broadcast Background Sound Balance Using Loudness for Elderly Listeners	560
<i>Tomoyasu Komori, Tohru Takagi, Koichi Kurozumi, Kiyotake Shoda, and Kazuhiro Murakawa</i>	
Automatic Live Monitoring of Communication Quality for Normal-Hearing and Hearing-Impaired Listeners	568
<i>Jan Rennies, Eugen Albertin, Stefan Goetze, and Jens-E. Appell</i>	
Portrait: Portraying Individuality	576
<i>Gemma Webster, Deborah I. Fels, Gary Gowans, and Norman Alm</i>	
Mental Models of Menu Structures in Diabetes Assistants	584
<i>André Calero Valdez, Martina Ziefle, Firat Alagöz, and Andreas Holzinger</i>	
Touch Screen User Interfaces for Older Subjects: Effect of the Targets Number and the Two Hands Use	592
<i>Guillaume Lepage and Nadine Vigouroux</i>	
Using a Wearable Insole Gait Analyzing System for Automated Mobility Assessment for Older People	600
<i>Johannes Oberzaucher, Harald Jagos, Christian Zödl, Walter Hlauschek, and Wolfgang Zagler</i>	
Author Index	605

Table of Contents – Part I

Design for Adaptive Content Processing

Design for Adaptive Content Processing: Introduction to the Special Thematic Session	1
<i>David Crombie</i>	
Generating DAISY Books from OpenOffice.org	5
<i>Christophe Strobbe, Jan Engelen, and Vincent Spiewak</i>	
Combining Web Services and DAISY for the Production and Delivery of Audio Tactile Diagrams	12
<i>Declan McMullen and Donal Fitzpatrick</i>	
Wiki, a New Way to Produce Accessible Documents	20
<i>Alex Bernier, Dominique Burger, and Bruno Marmol</i>	
Guided Generation and Evaluation of Accessible Scalable Vector Graphics	27
<i>Bernhard Dürnegger, Christina Feilmayr, and Wolfram Wöß</i>	
Improving the Re-digitisation Process by Using Software with Automatic Metadata Detection	35
<i>Reinhard Ruemer, Klaus Miesenberger, Franz Kummer, and Claus Gravenhorst</i>	
New Production and Delivery System for Pupils with Disabilities in Austria as Chance for Higher Quality Output	43
<i>Reinhard Ruemer and Klaus Miesenberger</i>	
A Flexible Software Architecture Concept for the Creation of Accessible PDF Documents	47
<i>Alireza Darvishy, Hans-Peter Hutter, Alexander Horvath, and Martin Dorigo</i>	

Digital Access to Documents for People with Print Disabilities

DAISY Kindergarten Books and Rocket Science Journals	53
<i>John A. Gardner, Carolyn K. Gardner, Vladimir Bulatov, Blake Jones, Elizabeth Jones, Robert A. Kelly, Masakazu Suzuki, and Katsuhito Yamaguchi</i>	

Exploiting the Daisy Format for Automated, Single-Source Braille Publishing	57
<i>Lars Ballieu Christensen</i>	
Rapid Listening of DAISY Digital Talking Books by Speech-Rate Conversion Technology for People with Visual Impairments	62
<i>Naoyuki Tazawa, Shinichi Totihara, Yukio Iwahana, Atsushi Imai, Nobumasa Seiyama, and Toru Takagi</i>	
E-Books and Audiobooks: What about Their Accessibility?	69
<i>Jan Engelen</i>	
E-Books and Inclusion: Dream Come True or Nightmare Unending?	74
<i>Alistair McNaught, Shirley Evans, and Simon Ball</i>	
Collaboratively Generated Content on the Audio-Tactile Map	78
<i>Armin B. Wagner</i>	
Generating Braille from OpenOffice.org	81
<i>Bert Frees, Christophe Strobbe, and Jan Engelen</i>	
Sonification of ASCII Circuit Diagrams	89
<i>Angela Constantinescu and Karin Müller</i>	
Accessing Google Docs via Screen Reader	92
<i>Maria Claudia Buzzi, Marina Buzzi, Barbara Leporini, Giulio Mori, and Victor M.R. Penichet</i>	
Making Digital Maps Accessible Using Vibrations	100
<i>Bernhard Schmitz and Thomas Ertl</i>	
Non-visual Navigation of Spreadsheet Tables	108
<i>Iyad Abu Doush and Enrico Pontelli</i>	
New Testing Method for the Dyslexic and the Newly Blind with a Digital Audio Player and Document Structure Diagrams	116
<i>Mamoru Fujiyoshi, Akio Fujiyoshi, and Toshiaki Aomatsu</i>	
Annotating and Describing Pictures – Applications in E-Learning and Accessibility of Graphics	124
<i>Ivan Kopeček and Radek Ošlejšek</i>	
New Concepts of PC Interaction for Blind Users Based on Braille Displays with ATC Technology (Active Tactile Control)	131
<i>Siegfried Kipke</i>	

HCI and Non Classical Interfaces

Inclusive E-Services for All: Identifying Accessibility Requirements for Upcoming Interaction Technologies	135
<i>Alejandro Rodriguez-Ascaso, Erik Zetterström, Martin Böcker, Helge Hüttenrauch, Michael Pluke, and Matthias Schneider</i>	
HAIL: Hierarchical Adaptive Interface Layout	139
<i>John Magee and Margrit Betke</i>	
Customizable Software Interface for Monitoring Applications	147
<i>Manuel Merino, Isabel Gómez, Octavio Rivera, and Alberto J. Molina</i>	
Towards to Real-Time System with Optimization Based Approach for EOG and Blinking Signals Separation for Human Computer Interaction	154
<i>Robert Krupiński and Przemysław Mazurek</i>	
Towards Accessible Interactions with Pervasive Interfaces, Based on Human Capabilities	162
<i>Matthew Tylee Atkinson, Yunqiu Li, Colin H.C. Machin, and David Sloan</i>	
Accelerometer and Spatial Audio Technology: Making Touch-Screen Mobile Devices Accessible	170
<i>Flaithri Neff, Tracey J. Mehigan, and Ian Pitt</i>	

HCI: Software Accessibility and Social Interaction

An Open Source / Freeware Assistive Technology Software Inventory . . .	178
<i>Alexandros Pino, Georgios Kouroupetroglou, Hernisa Kacorri, Anna Sarantidou, and Dimitris Spiliotopoulos</i>	
Designing and Developing Accessible Java Swing Applications	186
<i>Theofanis Oikonomou, Konstantinos Votis, Dimitrios Tzouvaras, and Peter Korn</i>	
Effects of Visual Stimuli on a Communication Assistive Method Using Sympathetic Skin Response	189
<i>Fumihiko Masuda and Chikamune Wada</i>	
Computerized Assessing the Mouse Proficiency through Multiple Indicators	193
<i>Ming-Chung Chen, Yun-Lung Lin, and Chien-Chuan Ko</i>	
PUIR: Parallel User Interface Rendering	200
<i>Kris Van Hees and Jan Engelen</i>	

Transient Cooperation in Social Applications for Accessibility Mapping 208
Harald Holone and Jo Herstad

A Social Approach to Accessible Social Networking Using the CAT Model 216
Jennifer George, Gilbert Cockton, and Thomas Greenough

Entertainment Software Accessibility

Entertainment Software Accessibility:
 Introduction to the Special Thematic Session 224
Dominique Archambault

User Interface Evaluation of Serious Games for Students with Intellectual Disability 227
Cecilia Sik Lanyi, David J. Brown, Penny Standen, Jacqueline Lewis, and Vilma Butkute

Making Mainstreaming Videogames More Accessible: A Pilot Study Applied to Buzz!™ Junior Monster Rumble for PlayStation 235
Elena Laudanna, Maria Bulgheroni, Francesca Caprino, and Serenella Besio

Accessibility of a Social Network Game 243
Roland Ossmann and Klaus Miesenberger

Using Placement and Name for Speaker Identification in Captioning 247
Quoc V. Vy and Deborah I. Fels

A Real-Time Network Board Game System Using Tactile and Auditory Senses for the Visually Impaired 255
Daisuke Nagasaka, Naoki Kaneda, Kazunori Itoh, Makoto Otani, Michio Shimizu, Masahiko Sugimoto, Masami Hashimoto, and Mizue Kayama

Accessible Tourism

Realizing Accessible Tourism by Matching Stakeholders:
 Introduction to the Special Thematic Session 263
Kerstin Matausch and Franz Pühretmair

Planning of Inclusive and Accessible Events 266
Kerstin Matausch and Klaus Miesenberger

User Feed-Back in the Development of an Information System for Public Transport 273
Christian Bühler, Helmut Heck, Christian Radek, Rainer Wallbruch, Josef Becker, and Claudia Bohner-Degrell

User-Specific Web-Based Route Planning	280
<i>Bettina Pressl, Christoph Mader, and Manfred Wieser</i>	

Smart and Assistive Environments

Results of a Workshop Series on Ambient Assisted Living	288
<i>Martin M. Morandell and Erwin Fugger</i>	

Challenges for Norwegian PC-Users with Parkinson’s Disease – A Survey	292
<i>Miriam Eileen Nes Begnum</i>	

Usability and Usefulness of GPS Based Localization Technology Used in Dementia Care	300
<i>Øystein Dale</i>	

Development of Universal Communication Aid for Emergency Using Motion Pictogram.....	308
<i>Mari Kakuta, Kaoru Nakazono, Yuji Nagashima, and Naotsune Hosono</i>	

Friendly Human-Machine Interaction in an Adapted Robotized Kitchen.....	312
<i>Joan Aranda, Manuel Vinagre, Enric X. Martín, Miquel Casamitjana, and Alicia Casals</i>	

Audio Classification Techniques in Home Environments for Elderly/Dependant People	320
<i>Héctor Lozano, Inmaculada Hernández, Artzai Picón, Javier Camarena, and Eva Navas</i>	

Autonomamente: Using Goal Attainment Scales to Evaluate the Impact of a Multimodal Domotic System to Support Autonomous Life of People with Cognitive Impairment	324
<i>Thimoty Barbieri, Piero Fraternali, Antonio Bianchi, and Clarissa Tacchella</i>	

PMD: Designing a Portable Medicine Dispenser for Persons Suffering from Alzheimer’s Disease.....	332
<i>Roel de Beer, Roy Keijers, Suleman Shahid, Abdullah Al Mahmud, and Omar Mubin</i>	

Usable Web Accessibility: Editing

Merging Web Accessibility and Usability by Patterns	336
<i>Helmut Vieritz, Daniel Schilberg, and Sabina Jeschke</i>	

Towards the Convergence of Web 2.0 and Semantic Web for E-Inclusion	343
<i>Laura Burzagli, Andrea Como, and Francesco Gabbanini</i>	
Beyond a Visuocentric Way of a Visual Web Search Clustering Engine: The Sonification of WhatsOnWeb	351
<i>Maria Laura Mele, Stefano Federici, Simone Borsci, and Giuseppe Liotta</i>	
Editing Web Presentations by Means of Dialogue	358
<i>Luděk Bártek</i>	
Fast Access to Web Pages with Finger Reading	366
<i>Toshihiro Kanahori</i>	
The Evaluations of Deletion-Based Method and Mixing-Based Method for Audio CAPTCHAs	368
<i>Takuya Nishimoto and Takayuki Watanabe</i>	
Integrating Semantic Web and Folksonomies to Improve E-Learning Accessibility	376
<i>Iyad Abu Doush and Enrico Pontelli</i>	
Simple Browser Construction Toolkit	384
<i>Osamu Miyamoto</i>	

Usable Web Accessibility: Education and Motivation

Organizational Motivations for Web Accessibility Implementation – A Case Study	392
<i>Marie-Luise Leitner and Christine Strauss</i>	
Using Collaborative Learning to Teach WCAG 2.0	400
<i>Fernando Alonso, José L. Fuertes, Ángel L. González, and Loïc Martínez</i>	
Web_Access: Education on Accessible Web Design	404
<i>Klaus Miesenberger, Barbara Hengstberger, and Mario Batusic</i>	

Usable Web Accessibility: Evaluation

Assessing WCAG 2.0 Conformance in Practice	408
<i>Sylvie Duchateau, Denis Boulay, and Dominique Burger</i>	
Is the Accessibility Audit Dead?	413
<i>Joshue O Connor</i>	

Evaluating Conformance to WCAG 2.0: Open Challenges	417
<i>Fernando Alonso, José L. Fuertes, Ángel L. González, and Loïc Martínez</i>	
Automatic Checking of Alternative Texts on Web Pages	425
<i>Morten Goodwin Olsen, Mikael Snaprud, and Annika Nietzio</i>	
OASES: Online Accessibility Self Evaluation Service - A Web-Based Tool for Education Institutions to Self-Assess the Accessibility of Their Practice	433
<i>Simon Ball, Alistair McNaught, Sue Watson, and Andrew Chandler</i>	
Analysis and Evaluation of the Accessibility to Visual Information in Web Pages	437
<i>Youssef Bou Issa, Mustapha Mojahid, Bernard Oriola, and Nadine Vigouroux</i>	
User Testing of Google Reader and RIA Complexity — A Warning	444
<i>Joshue O Connor</i>	
Analyzing Effect of Web Accessibility — A Framework to Determine Changes in Website Traffic and Success	449
<i>Rudolf Hartjes and Christine Strauss</i>	
eGovernment Accessibility	
Implementation Concept for an Accessible Web CMS	456
<i>Dietmar Nedbal and Gerald Petz</i>	
Requirements of the Vision Impairments for E-Government Services in Taiwan from Accessibility to Efficiency	464
<i>Chi Nung Chu</i>	
Accessibility of eGovernment Web Sites: Towards a Collaborative Retrofitting Approach	468
<i>Annika Nietzio, Morten Goodwin Olsen, Mandana Eibegger, and Mikael Snaprud</i>	
Improving the Accessibility of Fabasoft Folio by Means of WAI-ARIA	476
<i>Mario Batusic and Daniela Steiner</i>	
Accessibility of Blended and E-Learning for Mature Age and Disabled Students and Staff	
Accessibility of Blended and E-Learning for Mature Age and Disabled Students and Staff: Introduction to the Special Thematic Session	484
<i>Helen Petrie, Christopher Power, Carlos A. Velasco, and Jesus G. Boticario</i>	

Understanding and Supporting the Needs of Educational Professionals Working with Students with Disabilities and Mature Age Students	486
<i>David Swallow, Helen Petrie, and Christopher Power</i>	
Institutional Conditions for the Implementation of Accessible Lifelong Learning (ALL) Based on the EU4ALL Approach	492
<i>Cäcilia Weiermair-Märki and Elisabeth Unterfrauner</i>	
Design Guidelines for Developing Curriculum-Focused ICT Materials for Diverse Students	495
<i>Cara Nicole Greene</i>	
A Framework to Support Development of Learning Applications for Disabled Children	503
<i>José L. Fuertes, Ángel L. González, Gonzalo Mariscal, and Carlos Ruiz</i>	
Using Robots in Education and Therapy Sessions for Children with Disabilities: Guidelines for Teachers and Rehabilitation Professionals . . .	511
<i>Francesca Caprino, Serenella Besio, and Elena Laudanna</i>	
Virtual Learning Environments: Another Barrier to Blended and E-Learning	519
<i>Christopher Power, Helen Petrie, Vasily Sakharov, and David Swallow</i>	

Successful Service Provision

ICT for Persons with Disabilities: Bridging the Digital Divide in Bangladesh	527
<i>Mahjabeen Khaled Hossain</i>	
Success through Exchange: The Higher Education Accessibility Guide (HEAG)	531
<i>Andrea Petz and Klaus Miesenberger</i>	
Linking Instruments and Documenting Decisions in Service Delivery Guided by an ICF-Based Tool for Assistive Technology Selection	537
<i>Emily Steel, Gert Jan Gelderblom, and Luc de Witte</i>	
Easy-to-Use Social Network Service	544
<i>Niina Sillanpää, Sami Älli, and Timo Övermark</i>	

Standards: A Driver for Accessibility and Usability

Standards Are Drivers for Accessibility and Usability: Introduction to the Special Thematic Session	550
<i>Jan Engelen and Christian Bühler</i>	

International AT and DfA Standardisation: What Is in the Pipeline? ... <i>Jan Engelen</i>	553
Redefining Assumptions: Accessibility and Its Stakeholders <i>Rui Lopes, Karel Van Isacker, and Luís Carriço</i>	561
Stand4All: Promoting More Accessible Standards through Training of Stakeholders <i>Christophe Strobbe, Charlotte Mosies, Christian Bühler, and Jan Engelen</i>	569
Standards-Based Content Resources: A Prerequisite for Content Integration and Content Interoperability <i>Christian Galinski and Karel Van Isacker</i>	573
Standardization and Sustainability of Open Source Products - Barriers and Conflicts Demonstrated with the Linux Screenreader SUE -..... <i>Andrea Gaal, Gerhard Jaworek, and Joachim Klaus</i>	580
The Challenge of Mainstreaming ICT Design for All <i>Gill Whitney, Suzette Keith, and Barbara Schmidt-Belz</i>	583
Evaluating the Users' Satisfaction Using Inclusive Initiatives in Two Different Environments: The University and a Research Conference..... <i>Ana Iglesias, Lourdes Moreno, Javier Jiménez, and Pablo Revuelta</i>	591
People with Speech Impairment: “Weak and Silent Speech”: Tools to Support People with Speech Impairment	
ICCHP Keynote: Recognizing Silent and Weak Speech Based on Electromyography <i>Tanja Schultz</i>	595
CanSpeak: A Customizable Speech Interface for People with Dysarthric Speech..... <i>Foad Hamidi, Melanie Baljko, Nigel Livingston, and Leo Spalteholz</i>	605
Adding Voice to Whisper Using a Simple Heuristic Algorithm Inferred from Empirical Observation <i>Rüdiger Heimgärtner</i>	613
LIPPS — A Virtual Teacher for Speechreading Based on a Dialog-Controlled Talking-Head..... <i>Hermann Gebert and Hans-Heinrich Bothe</i>	621

People with Speech Learning and Cognitive Problems: Easy-to-Web

Easy-to-Web:

Introduction to the Special Thematic Session	630
<i>Cordula Edler and Birgit Peböck</i>	
The Need for Easy-to-Read Information on Web Sites	634
<i>Ulla Bohman</i>	
EasyWeb – A Study How People with Specific Learning Difficulties Can Be Supported on Using the Internet	641
<i>Kerstin Matausch and Birgit Peböck</i>	
Supporting the Web Experience of Young People with Learning Disabilities	649
<i>Harald Weber and Cordula Edler</i>	
Easy-to-Web Search for People with Learning Disabilities as Part of an Integrated Concept of Cognitive Web Accessibility	657
<i>Stephan Bergmann and Markus Erle</i>	
Adaptive Reading: A Design of Reading Browser with Dynamic Alternative Text Multimedia Dictionaries for the Text Reading Difficulty Readers	661
<i>Chi Nung Chu and Yao-Ming Yeh</i>	
In-Folio: An Open Source Portfolio for Students with Learning Disabilities	665
<i>Simon Ball, Andrew Chandler, Emma Millard, and John Sewell</i>	
Author Index	669

Methodological Considerations for Involving SpLD Practitioners and Specialists in Designing Interactive Learning Systems

Latifa Al-Abdulkarim, Areej Al-Wabil, Maha Al-Yahya,
Abeer Al-Humaimedy, and Sara Al-Khudair

Department of Information Technology, College of Computer and Information Sciences,
King Saud University, Riyadh, Saudi Arabia
{lalabdulkarim,aalwabil.c,malyahya,abeer,salkhudair}@ksu.edu.sa

Abstract. User involvement in designing learning environments to support individuals with Specific Learning Difficulties (SpLDs) is essential, particularly in inadequately examined languages such as Arabic. Three interactive systems to support students with SpLDs, two for students with dyslexia and one for students with dyscalculia were developed in a design-based research approach. In this paper, we describe a number of user involvement issues that emerged in the context of developing interactive learning systems for children with SpLDs in Arabic-speaking target populations. Findings indicate that language, context and culture emerge as challenges in creative and exploratory design approaches. Some of the ways these challenges have been approached are outlined.

Keywords: SpLD, Dyslexia, Arabic Software.

1 Introduction

Design is often seen at the heart of software engineering, a practice directed at the production of software systems. While research examining user involvement in designing systems for people with special needs has yielded insights into methodological considerations, work on involving users with SpLDs has often focused on design implications of tools or applications [1-3]. User involvement issues in designing systems to support the learning of people with SpLDs are often ignored or relegated to practicing software engineers in industry. This study sought to explore how involving SpLD practitioners and specialists in the context of Arabic-speaking populations can aid in the design of software that would be accommodating to varying learner needs. The specific research questions posed in this study were: How can the prior knowledge and experience of SpLD specialists be considered in the software design process? And how can software designers elicit requirements effectively? In this paper, we present a summary of the acquired experience in involving specialists and practitioners in the design process of interactive learning systems for people with SpLDs, particularly Arabic-language users, as an encouragement to others considering engaging in this challenging but rewarding process.

2 Method

For this study, we used a design-based research approach [4] to devise three interactive learning systems. These systems were developed as learning support tools through which students with SpLDs acquire skills, usually with the help of SpLDs practitioners in what is often referred to as ‘resource rooms’ in the local context, specific to reading in the Arabic language or math.

Procedure. Two SpLD practitioners and one domain specialist in SpLDs were involved, one for each system proposed. Sixteen groups of software engineering students participated as designers of bespoke software systems; each group was comprised of 5 to 7 developers. The study was conducted for a period of four months commencing with requirements discovery and ending with system deployment. Three requirements’ discovery sessions were conducted; each involved the participation of designers, researchers and the SpLD specialists to elicit the systems’ functional and design requirements. Each session was composed of two parts: First, SpLD specialists described their instructional design models and demonstrated their handmade artifacts and tools used in resource rooms that have been shown to be effective in their field. Following that, designers explored different ways of transferring these strategies and artifacts into interactive systems by proposing ideas and eliciting the specialists’ views. Designers were also offered the opportunity to include innovative interactive games supporting the teaching strategy. Iterative design allowed for some prototypes to be reviewed by specialists and refined before deployment. At the end of the project, specialists were involved in evaluating the functionality, accessibility, and usability of the systems.

3 Results and Discussion

Three types of software systems were developed. Two programs focused on Dyslexia in Arabic and one program focused on Dyscalculia with Eastern Arabic numerals. These are described in turn (Fig. 1 shows an example of these interactive systems).




Interactive Systems		
Dyslexia		Dyscalculia
		
Arabic prefix Al & silent consonants	Arabic long & short vowels (diacritics)	Math using Eastern Arabic numerals
7 systems, 35 designers	4 systems, 27 designers	5 systems, 30 designers

Fig. 1. Types of interactive systems developed for Arabic-speaking children with Dyslexia and Dyscalculia

Table 1. Mapping user requirements to design elements in the software

User Requirement	Design Element
Artifacts were demonstrated to designers in order to transform hands-on interaction (e.g. placing an element on a physical tree object) to an activity in the learning system.	Games were designed with click-drag & drop ways of interaction to simulate the activity with immediate visual indications reflecting the placement.
Specialists described lessons and scenarios that they use to support learning of concepts to students (e.g. stories).	Transforming these stories and lessons into instructional multimedia videos.
Considerations in how feedback is presented to students, especially when incorrect responses are exhibited.	Feedback regarding incorrect answers was positive and included visual cheerful characters to encourage users to try again.
Avoid animations as they distract students from the task at hand.	Animations were not incorporated in the system. Static images of characters were used and animated text was avoided.
Specific considerations in how text is read out-loud were demonstrated by a reading from the SpLD practitioner that emphasizes clear pronunciation, appropriate speed and level of pitch. Following text with a reading pacer (e.g. finger, ruler) was relevant to some parts of the system.	Spoken text in the system was re-created by designers based on the demo by practitioners. Synchronization of highlighting with spoken text was included to represent reading pacers to keep users' attention on the text being read.
The system needed to be used by students without the need for assistance from specialists.	Usability was a key consideration. Instructions were offered visually and in audio format as well as including videos demonstrating help for specific tasks.

3.1 Interactive Learning Systems

The first program was the Raindrop Program for learning the 'Al' prefix for dyslexic students. This program was to teach students how to discriminate between the 'Al' in the two formats, either silent or spoken in Arabic texts depending on the context of use. The second program was the Long & Short Vowel program for dyslexics. This program was designed to teach students with dyslexia how to discriminate between long vowels and short vowels as they are represented as diacritics in the Arabic language. The third program was the Basic Maths program using Arabic numerals for students with dyscalculia. This program focused on basic skills of addition and subtraction of 1-2 digit numbers in accordance with the 1st grade curriculum.

3.2 Understanding User Requirements

Creative solutions. Creativity was exhibited by designers in modifying the simple tutorials and lessons that specialists presented in the requirements discovery sessions. Some ideas were presented to specialists for feedback before further implementation was considered.

Cultural and language considerations. Selection of characters was a key design consideration as the appearance and name need to be extracted from the local context. This is essential for child engagement in the interactive learning system. Furthermore, designers initially did not consider exact matching of spoken feedback to the text appearing in the interface, which later on was highlighted by practitioners as an important design element so that students with auditory or language processing difficulties receive similar responses from the system. Another issue that emerged was considering grammar rules of spoken language in auditory feedback; this was a challenge for designers due to the complexity of the Modern Standard Language (MSA) which is used in teaching that differs from the spoken Arabic used in general populations.

4 Conclusion

This study has shown that the involvement of SpLD specialists in the design process is essential for efficient transfer of domain knowledge, and that creativity and exploratory approaches can be effective when language, context and culture are taken into consideration. Furthermore, the design based research method has been effective in advancing the design, research and practice of supporting the learning of students with SpLDs.

References

1. Dickinson, A., Eisma, R., Gregor, P.: Challenging Interfaces / Redesigning users. In: Proceedings of CUU 2003, Vancouver, British Columbia (November 2003)
2. Gregor, P., Newell, A.: An empirical investigation of ways in which some of the problems encountered by some dyslexics be alleviated using computer techniques. In: Proceedings of Assets 2000, pp. 85–91 (2000)
3. Petrie, H., Weber, G., Fisher, W.: Personalisation, interaction, and navigation in rich multimedia documents for print-disabled users. *IBM Systems Journal* 44(3), 629–635 (2005)
4. Wang, F., Hannafin, M.J.: Design-based research and technology-enhanced learning environments. *Educational Technology Research and Development* 53(4), 5–23 (2005)

Involving Users in the Design of ICT Aimed to Improve Education, Work, and Leisure for Users with Intellectual Disabilities

Emanuela Mazzone¹, Emmanuelle Gutiérrez y Restrepo¹, Carmen Barrera¹, Cecile Finat¹, Olga C. Santos¹, Jesús G. Boticario¹, Javier Moranchel², Jose Ramón Roldán³, and Roberto Casas²

¹ aDeNu Research Group, UNED

² University of Zaragoza

³ ATADES Association, Spain

{emazzone, egutierrez, carmen.barrera, cecile.finat, ocsantos, jgb}@dia.uned.es, {javiermo, rcasas}@unizar.es, joseramon.roldan@atades.org

Abstract. In this paper we describe different research projects involving users with intellectual disabilities. All these projects aim to enhance daily life activities and make their social integration possible, as the use of interactive technologies plays a significant role in supporting their independence. The users' participation throughout the design process was essential for achieving a usable and accessible product. We conclude by underlying the importance of adopting a user centred design methodology for the specific user group and the potential of an adaptive system to improve their learning experience.

Keywords: Accessibility, Adaptability, Usability, e-learning, independent life.

1 Introduction

After many years of research and practice regarding the interaction between people and machines, it is commonly agreed that user involvement is essential during the design and the development phases of novel applications and services. When users are drawn in, the solutions developed are expected to be more practical and address concrete user needs [10]. Nevertheless, developers and users are very often disconnected from each other. This calls for research on how to integrate all users (each person involved in the scenario, from carers to public agents, including also family and fellowships) throughout the design lifecycle. This approach is especially relevant in projects addressing users with special needs, where there is still a need for further analysis and for exchanging different findings.

Research in advanced learning technologies has a positive impact on the quality of life of disabled people since it can provide them with the means to increase their personal autonomy. In this sense, the CISVI (TSI-020301-2009-20), AMI4Inclusion (TSI-020100-2008-322) and ACCEDO2.0 (TSI-040200-2009-137) [2] projects, which we present in this paper, share a common challenge, helping people with intellectual disabilities to easily undertake daily life activities. In these projects, we focus

on the accessibility, usability, e-learning, e-mentoring and user modelling through diverse implementations, developments and evaluations by following a design approach that focuses on the users. However, these projects address different scenarios: from addressing basic activities for independent life [3] to training for insertion in the workplace [13], and from enhancing the learning experience in an educational context to creating a safe and accessible social network for recreation.

In this document, when referring to disability, we mainly focus on intellectual and developmental dysfunctions that affect an individual's social, conceptual and practical skills. Most people with an intellectual disability, if diagnosed on time, can have the opportunity to achieve a sufficient level of independence in everyday life. However research towards supporting independent living for people with intellectual disabilities is not so extensive. Within this target group, user needs vary widely, and there is a very broad range of e-Accessibility and usability challenges to be addressed.

2 State of the Art

Technology offers new ways to overcome specific problems associated with intellectual disability: it allows tutors to produce effective worksheets and provide an organised environment, improving accuracy, presentation and thus motivation and self-esteem [4]. However, specific intervention programmes have only been developed in the last 10 years (e.g. Hutinger, et al., [5]), making it difficult to assess these results and the real capacity of people with intellectual disabilities. According to the developers of online learning modules and games for individuals with Down syndrome [8] "taking the time to value and understand design from the user's perspective can chart unknown territory and yield transformational results". In the same line, several companies have used interactive tools to support workers with special needs: task organization and communication devices are used as task reminders (Highmark in USA, JobMatch in Australia), provide instructions (Cincinnati Children's Hospital) or communication through images [1].

The research at the Rix Centre [12] showcases different projects that highlight the potential of multimedia and ICT to enhance daily life. They promote different ways in which people with learning disabilities can explore their personal options and make informed choices, tell their own stories, using interactive media. They also propose effective strategies for encouraging designers to produce websites that are accessible to people with learning disabilities. An example of their work in this direction is a web portal designed specifically with the intended users in mind. The findings from the evaluation with users showed evidence of participant self-directed interest and initiated use of web technologies [17]. A similar study [14] explored the development of an alternative web browser for individuals with cognitive disabilities. They observed how participants interacted with different versions of the browser in order to find the optimal way to enhance their browsing experience.

Still far from being a universally adopted approach, many experimental projects like these are starting to consider final users when developing technology intended for people with intellectual disabilities. The majority of them do so mainly in the final evaluation stages, while others involve them also in the design phase. We adopted this latter approach, as we consider it has a greater potential for achieving usable and accessible products.

3 The Methodological Approach

According to ISO 13407 [6], before undertaking the design process, designers need to understand and specify who the intended users are, what their goals and tasks are and what is their context of use. The success of a product is based on this understanding, where its usability determines the effectiveness, efficiency and satisfaction with which specified users achieve specified goals in a particular environment [7]. The user-centred design (UCD) approach places the user as the core of the solution's development process and establishes that "the needs of the users should dominate the design of the system" [10]. Adopting a UCD approach helps to define the use of the product in the context of a real situation, based not solely on the designer's knowledge or appropriate design guidelines, but on inquiries in the field that lead to identify the features that are really necessary. This process implies iterative design and evaluation cycles, where initial ideas are prototyped and tested in context to identify possible improvements.

When applying the methodology to our design contexts, we needed to address the users' different needs and practices which depended both on their individual characteristics and on the approach taken by the different centres involved. A thorough understanding of the users' profile and the product context of use helped to define functionalities and interactions for a suitable solution. The educators and professionals responsible for the user's activities we were designing for represented our primary source of information on the users and their context of use. As explained in the 'Discussion' section, field studies and methods were adapted and limited to the many constraints which emerged during the project. Initial information was obtained from the professionals and from observing the users' activity in place focussing on: i) the specific activities as well as the general behaviour in the centres; ii) the user abilities, their potential and limitations; iii) tools they were familiar with or not.

With this initial knowledge, having the product specifically defined for each context, it was possible to outline a first design of the interaction and its interface and start adaptation to both to the users' needs and preferences and to the technology and accessibility constraints. After several adjustments, validated with the educators and professionals of the centres, we tested a first working prototype with the final users. We paid particular attention to how easy was it for them to i) interpret and manage the functionalities; ii) navigate the structure of the flow of actions for the users; iii) understand the content; and, iv) possible technical flaws.

Depending on the peculiarities of each project, at least one cycle of the iteration design-evaluation was completed, with the results of the evaluation feeding improvements in the prototype. The final product is intended to be sustainable, so that at the end of the project each centre is able to use it and adapt its content to emerging activities and needs. Below, we describe the process carried out in three contexts of use: basic daily activities (in a home-like environment), professional tasks (in the workplace), and educational training (in the classroom).

4 R&D Work and Results

4.1 Daily Life Activities in Home-Like Environment

Within the CISVI project there are several Social Spaces for Innovation or ESdI related to e-Inclusion, e-training and e-health [3]. In one of the ESdI we try to encourage

independent living by learning daily life activities (ADL). This ESdI is developed within the ATADES Association; one of its main objectives is coaching people with intellectual disabilities in ADL. So far, such learning is based on templates and hand-made cards, with adapted pictograms. The new approach to learning is to adapt this methodology into technological support, as structured ADL learning units, via web (through universal design standards, usability and accessibility). The first design iteration involved using ICT for ADL support in the context of a mock apartment that ATADES uses for ADL training (using traditional methods). Here, six users with cognitive disability and a medium level of autonomy, together with 3 educators and 3 technologists were involved. We focussed on the ADL "Washing your hands" and adapted the methodology used by ATADES incorporating technological support: i) simple and sequential ADL steps; ii) different formats for the content including adapted text and audio with verbalizations, graphical information with pictures and adapted pictograms, browser buttons; iii) browsing support based on e-learning with multimodal interfaces (i.e. touch screen, adapted mouse, switches, etc.).

A worksheet was designed to assess the performance of the users. The activity was divided in sequential steps with different operational objectives, i.e. the successful execution of each step with and without verbal support. At the same time, we also verified the accomplishment of all the steps and registered the user's attitude towards the use of the device together with other observations. At the beginning, users learnt the ADL activity from the educators verbally, without technology, and at the end of the cycle, the ADL activity was learnt by users with technological support. Users paid attention to the system, in most cases their performance in certain steps improved. The response of some users was conditioned by the system's behaviour but occasional technological faults, acceptance and dexterity in the use of the technology (touch pad, navigation, etc.) improved during the evaluation. In general, s added a motivating factor for the acquisition of new knowledge and skills. As the users became familiar with the interface and learnt the steps to carry out the ADL, they also felt motivated to continue learning to enhance their autonomy.

4.2 Professional Tasks in the Workplace

In the same project (CISVI), another ESdI focussed on providing technological support for people with intellectual disabilities in their workplace [13]. It was developed within an educational programme (Promentor Programme) that provides classroom training for these future workers as well as continuous coaching during the working period. Personal coaches are in charge of contacting the company and adapting the workspace and tasks to the needs of the specific worker. They also provide workers with tailored tools (notes, guides, and memorandum) which provide support during their job activities.

The task of filing different types of documents was selected to be supported by a notebook. A thorough design process was needed in order to adjust the content of the currently used paper-based-support to the technological medium and to exploit its interactive functionalities. A first proposal of the conceptual model for the netbook interface was discussed together with the coacher. Then, we refined the task flow, colour scheme, navigation, labels, and icons to suit the worker's skills and needs.

To test the prototype with the worker in a controlled environment, we introduced her to the tool and let her navigate the interface freely and asked her to complete specific tasks. Apart from inviting her to comment on anything she liked, disliked or felt was missing, at the end of each task we also asked her to evaluate specific features of the interface. In this way, we were able to identify flaws and slips and adjust the prototype to her real preferences and requirements. After an initial wariness, the user started gaining confidence and familiarity with the tool and was able to appreciate its innovative functionalities but also expressed concerns and suggestions for the following implementation to evaluate in the context of the workplace.

4.3 Educational Training in the Classroom

The project AMI4INCLUSION is within the same Promentor programme. The context is a two-year Higher Education course “Training for the Inclusion of Young People with Intellectual Disabilities in Working Environments” [11]. In this course, groups of 15 students with intellectual disabilities, aged between 18 and 30, are trained to enter working environments. The main purpose of the project is to develop an e-training system that support workers with intellectual disabilities throughout their professional development and enhance their inclusion.

We focussed on the organising tasks and filing module, one of the most common tasks of their future jobs. At the beginning of the design process, we were able to observe the teacher and the students during a class and have an insight of the context, its challenges and potential. The groups are formed by a wide range of abilities and disabilities; which demands a high level of flexibility and adaptation. Despite differences in individual skills, what is common in the whole group is their high motivation to learn and achieve objectives. This attitude increases when using computers or technology devices, and makes them make an effort to concentrate and face the problems.

Through subsequent consultancy with teachers and coaches, we adapted the module’s theoretical notions in a linear learning unit with additional interactive support. Together with the learning unit, we implemented training exercises to rehearse the concepts learnt, while also exploring the integration of a recommender system to facilitate users’ learning experience. The tool needed to provide professional and adult-like graphic, simple instructions; visual consistencies; positive language; immediate visual and audio feedback.

The first prototype was tested during a class. The interactive exercises were engaging for everyone, despite their success in completion. Together with our observations and the teachers’ comments, the students told us about their preferences, their difficulties, and suggestions for improvement. All the information gathered from the first evaluation filtered through usability and accessibility standards, fed the second design cycle, where additional features were also included in the system and tested again. Besides its educational value and support for delivering varied and tailor-made exercises, the teachers noted its motivational power.

5 Discussions

These different case studies represent an interesting base for discussing the application of UCD methods in different design contexts to achieve an effective and accessible

product for users with intellectual disabilities. For the first case study the interaction needed to be as simple and as detailed as possible without any space for ambiguity or interpretations (i.e. shortcuts could have enabled a more agile interaction but were not suitable for the specific context of use). In the second project the product needed to be completely tailor-made to the individual user. The design followed a structure that allowed modularity and adaptability to different requirements, but the final product had to be customised for each user. Finally, in the last study, the solution needed to support the needs of different users with varied profiles and abilities.

The solutions proposed for each context sprung from the successful involvement of the users throughout the design process. The information needed for understanding the context of use and user requirements was not always easy to access, was and had to be sought indirectly. In the second case study, for example, it was not possible to enter the workplace and observe their typical working day. In that case, detailed descriptions of the working procedures, significant accounts, pictures of the different working spaces, were all essential for gaining enough knowledge to design effective solutions.

The role of the tutor and educator was crucial in all the studies. The person in charge of the activity provided information on the users and the context of use; this would have hardly been accessible otherwise. Direct contact with users with intellectual disabilities also required special preparation: psycho-educational expertise is required both to facilitate the communication with the users and to interpret their answers, when evaluating interactions and collecting feedback.

Another important aspect that emerged from the studies is the potential for adaptability. Educators and professionals can usually recognise the varying needs of the users and immediately adjust their training practice to the specific situation. Although far from substituting human presence, it is important to design the technology so that it can dynamically respond to the contextual needs of the users based on determined conditions and the users' interaction. This adaptive system is also useful for the tutor, providing real-time information on user interaction and allowing educators to make the appropriate content and methodology adjustments.

6 Conclusions and Future Work

We consider that these studies and their results provide a useful example of how design based on direct users' involvement is beneficial for the effectiveness of the product for users with intellectual disabilities. Similarly, we have shown the importance of having a solid understanding of usability and accessibility, as well as the strategies of Internet use for people with disabilities.

In each project, the design approach had to be adapted to the specific users and the context. From the experience in these three studies we can generalise some points that we consider essential when designing usable and accessible solutions for users with intellectual disabilities, and that, to some extent, are also beneficial in any design. Amongst all the findings, we mention just a few: i) the contact with the users has to be supported by a psycho-educational professional; ii) features that could be considered usable according to commonly applied principles need to be adjusted and tested in each specific context of use; iii) the WAI Guidelines [15,16] and personalisation

standards [7,8,9] implementation are essential for achieving the necessary flexibility and adaptability; iv) users' engagement is important for increasing the users' motivation for completing the task and for their familiarisation and satisfaction with using the product; vi) It is essential the participation of real experts in usability and accessibility, to filter the perceptions expressed by users and even by other researchers; vii) the system's adaptability and flexibility is crucial as the specific user requirements are changeable and very dependant both on unpredictable external and internal factors - more than with most users. We also found particularly effective the use of the recommender system [13] to provide content tailored to the needs of this group of users. Because users have changing needs, the recommender system can adjust the content and learning experience to the user's needs.

The expertise acquired with these previous studies is now being applied in the ongoing project ACCEDO 2.0. As mentioned earlier, the aim is to provide a safe and accessible social network for users with intellectual disabilities, where they can easily and safely share their ideas, news and files. This project shares the same design approach centred in the real needs of the users, where field studies and direct interaction with the users will guide the directions of the design.

Acknowledgements

We thank all the participants who welcomed us in their centres, providing valuable information and a remarkable experience. Authors also thank the Spanish government for supporting the projects, in particular AMI4INCLUSION (TSI-020100-2008-322), CISVI (TSI-020301-2009-20), ACCEDO 2.0 (TSI-040200-2009-137), and A2UN@project (TIN2008-06862-C04-01/TSI).

References

1. AbleLink Technologies Web Trek Connect, Axistive 2.0. - 2001 (April 12, 2010), <http://www.axistive.com/web-trek-connect.html>
2. aDeNu aDeNu Projects: Accedo 2.0; AMI4INCLUSION; CISVI: <http://adenu.ia.uned.es/web/en/projects>
3. Barrera, C., et al.: Empowering disabled people: e-learning challenges for independent life. In: TUMAS-A 2009 held in conjunction with AIED, Brighton, UK, pp. 38–44 (2009)
4. Black Bob Information - Articles: Down's Syndrome, Computers and the Curriculum [Online] Inclusive Technology (2006), <http://www.inclusive.co.uk/infosite/bblack.shtml>
5. Hutinger, P., Rippey, R., Johanson, J.: Findings of Research Study on Effectiveness of a Comprehensive Technology System Demonstrate Benefits for Children and Teachers. Western Illinois University (1999), <http://www.wiu.edu/thecenter/articles/findings.html>
6. ISO 13407:1999 Human-centred design processes for interactive systems
7. ISO/IEC 24751-1:2008 Information technology. Individualized adaptability and accessibility in e-learning, education and training. Part 1: Framework and reference model
8. ISO/IEC 24751-2:2008 Information technology – Individualized adaptability and accessibility in e-learning, education and training – Part 2: "Access for all" personal needs and preferences for digital delivery

9. ISO/IEC 24751-3:2008 Information technology – Individualized adaptability and accessibility in e-learning, education and training – Part 3: "Access for all" digital resource description
10. Norman D.A., Draper S.W.: User Centered System Design. New Perspectives on Human-Computer Interaction - (s.l.), p. 526. L. Erlbaum Associates Inc, Mahwah (1986)
11. PRODIS [Online] Fundación Prodis (2010), <http://www.fundacionprodis.org/>
12. Rix Centre. The Big Tree (2007), <http://thebigtree.org/>
13. Santos, O.C., Couchet, J., Boticario, J.G.: Personalized e-learning and e-mentoring through user modelling and dynamic recommendations for the inclusion of disabled at work and education. In: ICALT 2009, Riga (2009)
14. Sevilla, J., et al.: Web accessibility for individuals with cognitive deficits: a comparative study between an existing commercial web and its cognitively accessible equivalent. ACM Transactions on Computer-Human Interaction, 1–23 (2007)
15. W3C-WAI WCAG 2.0. (December 2008), <http://www.w3.org/TR/WCAG20/>
16. W3C-WAI Web Content Accessibility Guidelines 1.0, <http://www.w3.org/TR/WAI-WEBCONTENT/>
17. Williams, P., Hanson-Baldauf, D.: Testing a web information portal for people with learning disabilities. Journal of Research in Special Educational Needs, 42–51 (2010)

PDA Software Aimed at Improving Workplace Adaptation for People with Cognitive Disabilities

Alberto Ferreras¹, Juan-Manuel Belda¹, Ricard Barberà¹, Rakel Poveda¹, Miguel Urra², Nuria García², Miguel Tito¹, and Marta Valero¹

¹ Instituto de Biomecánica de Valencia,

Universidad Politécnica de Valencia, Valencia 46022, SP

² AFEM: Asociación FEAPS para el empleo, Madrid 28050, SP

Abstract. This article presents the objectives, methodology and results of a project aimed at improving personal autonomy and adaptation to the workplace of people with cognitive disabilities. These people may have difficulties to be independent in various areas of life, especially at the workplace. Many of these problems are related to issues such as time control, independence in performing tasks, work habits, interpersonal communication, etc. To promote autonomy in these areas, the tool GTT (Time and Task Manager) has been developed. This is a software application running on a mobile device (a mobile phone or PDA) aiming to ease the adaptation to the workplace of people with cognitive disabilities through tutoring tasks, major event management and strengthening of key aspects of work.

Keywords: Workplace adaptation, personal autonomy, software tool, cognitive disabilities, Personal Digital Assistant (PDA), mobile phone.

1 Introduction and State of the Art

The autonomy of people with disabilities is a key factor to ensure their integration into society. One way to maximize autonomy is work, and therefore, the labor integration is considered to be highly related with welfare. Achieving and maintaining employment enables benefits to the individual such as self-esteem and independence [4,5].

Within the group of people with disabilities, those suffering from cognitive limitations (mild or moderate) often have additional difficulties to get and keep a job, either in the regular market or in the sheltered employment. Some of these difficulties are related to issues such as *difficulty in controlling time, lack of independence in performing tasks, incorrect work habits or poor interpersonal communication* [7,11].

The use of new technologies in supporting people with cognitive disabilities is a resource used in recent years in various studies and by different entities. New technologies, particularly in the field of computer science, have proven to

be useful in areas such as *cognitive rehabilitation, sensory stimulation, learning literacy, alternative communication, adaptation to environments and situations with the use of virtual reality*, etc [9,10,1].

In this sense, portable devices such as PDAs and mobile phones are tools with a high potential for supporting people with cognitive disabilities, for example, to promote the reminiscence of events and activities and to carry out therapeutic and rehabilitative tasks. The small size of these devices and their versatility make them ideal for use in different areas of life, including work [13,12,10].

However, the standard software and existing commercial applications for PDAs and similar devices are generally not designed for that people with intellectual disabilities can use them easily, as they present complex environments and imply a high knowledge in the use of software. Several studies [3,16,15] have identified the main parameters that must be taken into account when designing appropriate software for people with cognitive disabilities: *clear screens, consistent commands and features from screen to screen, appropriate sequencing and pacing, frequent feedback, combined use of pictures and audio prompts for navigation, oversized buttons that maximize the effectiveness of touch screen, linear designs of software operation, error minimization features or customization options that allow professionals to adapt the software to the unique needs of a variety of end-users*.

In absence of conventional software adapted to these requirements, it is often necessary to design specific programs and environments for people with cognitive disabilities suited to their characteristics and needs. Some examples of programs that are designed with these objectives are the following:

- *PEATTM* [14]: a handheld computer system that provides cueing and scheduling assistance for individuals with memory, attention, and cognitive disorders related to a brain injury.
- *MAPS* (Memory Aiding Prompting System) [2]: prompting system on a PDA platform for individuals with cognitive disabilities with an interface for caregivers.
- *Comunicador CPA* [19]: Symbolic Communicator for PDA, that also allows to generate routines and notices.
- *SC@UT* [17]: Augmentative and alternative communication system that help people with autism and other communication problems.
- *TUTOR Project* [18]: a tutoring system, that runs on a mobile phone and its aim is to increase the autonomy of people with cognitive disabilities in labour and daily life activities.
- *AZAHAR* [20]: a set of applications installed on a PDA, including a simplified music player, a communicator, a graphic SMS editor and a tool to understand the concept of time.
- *AbleLink Handheld Solutions* [21]: Commercial tool for people who need support in their daily activities by providing step-by-step task prompting and scheduling support, combined with an easy to use desktop shell.

Most of these applications are directed toward areas related to ADL, rehabilitation, communication, location or spatial orientation. However, there are few

applications aimed at the labour adaptation. At the workplace can be very useful products that provide support in areas such as learning tasks, acquiring job skills, risk prevention, etc.

It is therefore very important to develop tools that make easier the adaptation to the workplace of people with intellectual disabilities and hence, enhance their personal autonomy and quality of life [8,18,6]. To do this, the Instituto de Biomecánica de Valencia (IBV) and the Asociación FEAPS para el empleo (AFEM), have developed a software tool named **GTT: Time and Task Manager**. GTT is a software application that can be implemented on a PDA or a tactile mobile phone, which aims to promote autonomy in the workplace for people with mild to moderate cognitive impairment. To do this, the software offers to the worker a guide through the different tasks to be performed, notifies important events and provides a strengthening of key aspects of work.

2 Methodology

To develop this application, a work plan divided into the following stages was followed:

- Document review, about difficulties and solutions for adapting the workplace and effects of technology in people with cognitive disabilities.
- Collect of information about tasks, problems and needs. To do this, several workplaces (gardening, cleaning and laundry) were selected and visited, performing focus groups and interviews with workers with intellectual disabilities and their supervisors.
- Tool architecture design and programming of a prototype.
- Evaluation of the prototype in the workplaces selected.
- Tool redesign, based on the information provided by employees and supervisors.
- Implementation and dissemination of the results.

3 Results

3.1 Prototype Evaluation

From the specifications obtained from the literature review, field study and discussion groups with technicians and supervisors, we made a first version of the software GTT. This version was implemented in two different portable devices: HP iPAQ 214 and ASUS P320 with the aim of testing for use in real situations. The tests were performed with 8 workers with mild cognitive impairment and 4 supervisors, belonging to 4 different workplaces.

This evaluation was carried out in a living lab context:

- Training of supervisors and workers on the use of software and the PDA.
- Living lab phase:

- Free use of software over a period of 2 months (1 month for each model of PDA) in real working conditions (See Figure 1).
- Sending of regular information about: mistakes, usability, design and suggestions for improvement.
- Panel discussion with supervisors and technicians in order to share the results of the evaluation process.



Fig. 1. Evaluation of the prototype by a gardening worker

The main results obtained in the evaluation phase were:

- Positive rating from both workers and supervisors in regard to the usefulness and usability of the application.
- Main issues and suggestions from supervisors:
 - Simplify the process of installing the application on the PDA.
 - Improve the usability of the management software.
 - Modify the design of the application: avoid the 'childlike appearance', but maintaining the simplicity of use, since it is an application for adults in work environments (See Figure2).
 - Correction of errors in the transcription of accents, size of characters and image display on the PDA.
- Main issues and suggestions from workers:
 - Preference of a large PDA (in this case the HP iPAQ 214, compared to the ASUS P320) because it is easier to manage the application with the fingers (no need to use the pointer) and the text and images are best viewed.
 - Modify the size and location of scroll arrows and buttons for easier handling.
 - Change the way of presenting the text in the alarms to avoid errors when combining the text with the time.

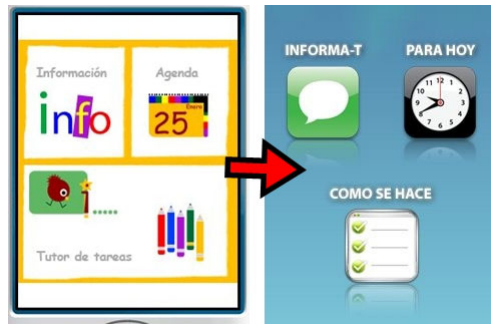


Fig. 2. Initial design of the PDA application and further modification based on user and supervisors preferences

3.2 GTT Software

Once the user testing was carried out, the final version of the software was developed. The software GTT is divided into two platforms:

- **Desktop application.** It is a PC program that helps to create and administer the content that will be subsequently provided to the worker.
- **Portable application** (See Figure 3). It is a software based on JAVA, running on a PDA or mobile tactile device, which uses the generated files in the desktop application. This is the application that will be used by the worker.

On both platforms, GTT has three specific sections:

- ‘*Cómo SE HACE*’ (*HOW IT IS DONE*). Provides a tutorial on the various tasks to be performed by the worker. The PDA shows, step by step, how to do a particular task. To do this, it uses a combination of text, images and audio. The administrator can design the structure of tasks and subtasks



Fig. 3. GTT – Screenshots of the portable application

more appropriate for an individual worker. This module can also be used as a personal communicator, composing messages that combine images and audio.

- ‘*PARA HOY*’ (*FOR TODAY*). This is a tailored agenda that includes major events that the worker has to remember in their workday. In the PDA the different items are available for a particular day. At the time of the event, an alarm notifies the employee of the event through a combination of text, images or audio.
- ‘*INFORMA-T*’ (*INFORM YOURSELF*). This module may contain additional activities (eg questionnaires) to reinforce important aspects of work: identification of places, equipment testing, prevention of occupational hazards, etc.

This structure is initially designed for the working environment, although applications may be broader: home environment, activities of daily living, alternative communication, etc. The PDA application is used by the disabled worker. Therefore, the interface design and the structure include a number of specific features:

- Interface with very simple structure and easy navigation.
- Big buttons that make it possible to handle both the finger as the pointer.
- Feasibility to combine text, images and audio according to user requirements.

4 Conclusions and Further Work

The software has been tested on various sheltered employment centres, to assess the usefulness of it, correct the existing defects and raise additional features for future projects. Concerning these aspects, the conclusions were as follows:

- Both workers and supervisors have valued highly the application. The flexibility of the software has allowed the adaptation to workers with different needs and difficulties. Supervisors have highlighted an improvement in the quality and speed of work and less need for training and consultations.
- It is important to develop materials, besides the software, adapted to both workers with cognitive disability and their supervisors. These materials must be useful not only to implement activities that may be used on GTT, but also to elaborate other tasks related with learning, training and sensitization. Some themes that are considered of high interest in order to develop materials are: *work risks prevention, ergonomic issues, use of machines and tools, reinforcement of proper behavior or money management*, among others.
- Finally, workers, supervisors and technicians have suggested ideas for increasing the performance of GTT: record of activities, checklists, music player, tutorials, etc. Many of these ideas will be developed throughout 2010.

Acknowledgements

The authors would like to thank to BBVA for its financial support of this project. Software development has been carried out by IVER. Special thanks to the

following sheltered employment centres, for their collaboration in the development of this project: APCOM ECOJARDİ̄ (Murcia), VEGAPRESAS (Asturias) and ASPROSERVIS-ASPRONA (Valencia).

References

1. Braddock, D., Rizzolo, M., Thompson, M., Bell, R.: Emerging technologies and cognitive disability. *Journal of Special Education Technology* 19(4), 49–53 (2004)
2. Carmien, S.: Maps: Dynamic scaffolding for independence for persons with cognitive impairments. In: Brusilovsky, P., Corbett, A.T., de Rosis, F. (eds.) *UM 2003. LNCS*, vol. 2702, pp. 408–410. Springer, Heidelberg (2003)
3. Davies, D.K., Stock, S.E., Wehmeyer, M.L.: 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 (2002)
4. Eggleton, I., Robertson, S., Ryan, J., Kober, R.: The impact of employment on the quality of life of people with an intellectual disability. *Journal of Vocational Rehabilitation* 13(2), 95–107 (1999)
5. Egado, I., Cerrillo, R., Carmina, A.: Social and labour inclusion of people with intellectual disability through supported employment programs. a challenge for guidance. *Revista española de orientación y psicopedagogía* 20(2), 135–146 (2009)
6. Eizmendi, G., Herreroand, M., González, M., Sagües, I.: New technologies to support the integration of people with disabilities in the workplace. In: Eizmendi, G., Azkoitia, J.M., Craddock, G. (eds.) *Challenges for Assistive Technology. AAATE 2007 Conference Proceedings*, pp. 131–135. IOS Press, Amsterdam (2007)
7. Ferreras, A., García, C., Tortosa, L., Prat, J.M., Molla, F., Navarro, J., Vera, P., Dejoz, R.: Ergonomic adaptation of workplaces for people with disabilities: methodology and examples. In: Eizmendi, G., Azkoitia, J.M., Craddock, G. (eds.) *Challenges for Assistive Technology. AAATE 2007 Conference Proceedings*, pp. 136–140. IOS Press, Amsterdam (2007)
8. Ferreras, A., Piedrabuena, A., Belda, J.M., Barberà, R., Oltra, A., Poveda, R., Prat, J., Tortosa, L.: Analysis and adaptation of workplaces for people with cognitive disabilities using software tools. In: Miesenberger, K., Klaus, J., Zagler, W.L., Karshmer, A.I. (eds.) *ICCHP 2008. LNCS*, vol. 5105, pp. 721–728. Springer, Heidelberg (2008)
9. Gentry, T.: Pda-enabled job coaching for individuals with acquired brain injury (2006), <http://www.vcu.edu/partnership/pda/Jobcoach/index.html>
10. Hart, T., Hawkey, K., Whyte, J.: Use of a portable voice organizer to remember therapy goals in traumatic brain injury rehabilitation: A within subjects trial. *Journal of Head Trauma Rehabilitation* 17(6), 556–570 (2002)
11. Ibáñez, P., Mudarra, M.J.: Integración sociolaboral: trabajadores con discapacidad intelectual en Centros Especiales de Empleo, Dykinson, Madrid (2007)
12. Kautz, H., Fox, D., Etzioni, O., Borriello, G., Arnstein, L.: An overview of the assisted cognition project. *Technical Report WS-02-02*, 60–65, AAAI (2002)
13. Lancioni, G.E., O'Reilly, M., Seedhouse, P., Furniss, F., Cunha, B.: Promoting independent task performance by persons with severe developmental disabilities through a new computer-aided system. *Behavior Modification* 24(5), 700–718 (2000)

14. Levinson, R.: The planning and execution assistant and trainer (peat). *Journal of Head Trauma Rehabilitation* 12(2), 85–91 (1997)
15. LoPresti, E.F., Mihailidis, A., Kirsch, N.: Assistive technology for cognitive rehabilitation: State of the art. *Neuropsychological Rehabilitation* 14, 5–39 (2004)
16. Okolo, C.M., Bahr, C., Reith, H.: A retrospective view of computer-based instruction. *Journal of Special Education Technology* 12, 1–27 (1993)
17. Paredes-Garrido, M., Rodríguez-Fórtiz, M., Medina-Medina, N.: Evolution of sc@ut adaptive augmentative communicators. In: Luca, J., Weippl, E. (eds.) *Proceedings of World Conference on Educational Multimedia, Hypermedia and Telecommunications*, VA, pp. 1112–1119. AACE, Chesapeake (2008)
18. Rubio, J., et al.: Tutor project: An intelligent tutoring system to improve cognitive disabled people integration. In: Miesenberger, K., Klaus, J., Zagler, W.L., Karshmer, A.I. (eds.) *ICCHP 2008. LNCS*, vol. 5105, pp. 729–732. Springer, Heidelberg (2008)
19. Velasco, R., Muñoz, D.: *Comunicador personal adaptable* (2003), <http://www.comunicadorcpa.com>
20. VVAA. *Proyecto azahar* (2009), <http://www.proyectoazahar.org>
21. VVAA. *Ablelink handheld solutions* (2010), <http://www.ablelinktech.com>

EasyICT: A Framework for Measuring ICT-Skills of People with Cognitive Disabilities

Jan Dekelver, Tim Vannuffelen, and Joan De Boeck

K-Point, K.H. Kempen University College
2440 Geel, Belgium

{jan.dekelver, tim.vannuffelen, joan.de.boeck}@khk.be

Abstract. Over the last decade, the basic skills to operate a computer (ICT skills) are an essential requirement to participate in the current digital era. People not possessing these skills are likely to miss access to education, entertainment, business and social life. In particular for people with cognitive limitations, this is a real threat. In the EasyICT project, we aimed at the development of a training and assessment framework, supported by an on-line tool. Using this tool youngsters with mental disabilities can be systematically tested for their ICT skills. As a result, they receive a report and additional training material in order to improve their skills.

1 Introduction

Over the last decade, ICT has become an essential part of our daily life. Hence, not being able to operate a computer takes away the chances to participate in this new digital society. This is in particular true for people with cognitive disabilities. Alternatively, making these new technologies available for those target groups, will contribute to full inclusion of these individuals in the contemporary society [1]. Several reports stress the impact on the social, economical and cultural development of different target groups when having internet connection *and* being able to use it [2][3]. Despite of the importance of possessing these ICT skills, there is almost no attention put focussing on methods for the acquisition and measuring of ICT skills by persons with cognitive disabilities. Hence, estimation of the needs is still purely empirical.

In our EasyICT project, we aim to address this problem. The goal is to formalise the process of training and assessment of ICT skills for the envisioned target groups. We focus on a very broad group of youngsters with cognitive disabilities between the age of 6 and 18. It is very important also to include those individuals who need full assistance to perform computer tasks.

Furthermore, in order to easily integrate the framework into existing educational systems, it is important that the realised framework fits into the different local European educational models and is compatible with the European Computer Driving Licence (ECDL) [4].

2 The Framework

In order to assess ICT skills, we categorised them in 4 main groups:

1. Managing the computer (identifying parts, handling mouse, ...)
2. Browsing the internet
3. Using e-mail
4. Using ICT to communicate in a safe, sensible and appropriate way

For each of these categories, the different aspects are identified. E.g. in the category ‘managing the computer’, one of the aspects is ‘hold the mouse steady’ or ‘click the mouse once’. Per aspect, several test questions are developed.

An assessment is always performed in one-to-one set-up with a tutor and one candidate. There is no time limit to complete an exercise or test. A test session, however, should not exceed one hour in order to avoid loss of concentration, but each test can be re-entered at any later time in case the candidate should lose their concentration.

Whenever possible, candidates are able to use their own familiar PC, but if for some practical reason a different PC has to be used, adaptations have to be made to come as close as possible to the familiar environment (system software, browser, adapted hardware, etc.).

In order to allow the test to bring a positive message to the candidate, the tutor will only select those tests that the candidate will most probably be able to answer. Immediately after the test, a certificate is provided for the candidate, listing all the skills that have been assessed positively. Keeping a positive message in mind, skills that have not been tested or achieved are not listed.

3 The Online Test Environment

3.1 Open Source Framework

Based upon the theoretical background and breakdown, acquired during the first part of this project, an online environment has been created. This environment is built upon a Dokeos e-learning platform [5]. The open source nature of this platform allowed us to easily add some additional functionality while removing other standard Dokeos functions. In order to give a reader an impression of the implementation issues, in what follows, we will shortly discuss some of those adaptations.

The final goal of the online test environment was to create a very easy to use tool for teachers in special education. Hence, all unnecessary functions were removed or disabled from the interface. Security of the children’s data is an important topic, but not at a cost of penalising usability. An issue rises specifically when a child moves to another class or changes to another tutor. The tool allows teachers to transfer children to their own class, but obviously we must prevent them from moving arbitrary students. The parents (or guardian) are considered as a trusted actor: in order to complete the transfer, the teacher needs a password. This password is printed on the certificate, which is handed over by the first tutor to the parents. They have to present the certificate with the password to the new teacher, who is instructed to change the password after the transfer is completed.

As stated before, another important requirement when dealing with children with intellectual disabilities, is that a tutor must be able to suspend a test at any arbitrary time when concentration of the candidate appears to fade away. Afterwards, at any later

time, one must be able to continue the test at the point where they left. Hence, when a test is restarted, previous results are automatically reloaded.

We are potentially dealing with an enormous number of candidates being tested; this was even the case in the pilot study that was running in five European countries. Therefore, in order to keep the number of students on the overview page reasonably small, teachers are able to deactivate one or more of their students. On the students overview page, only the picture of active students is shown. This means that a test can only be started for active students. A student, after being deactivated, can always be activated again. Typically, students that no longer need to perform the test are deactivated by their teacher.

Next, in the original Dokeos environment, scores were only shown numerically. Tailoring the interface more to the living environment of our target group, an extension has been made to visualise scores also in a graphical way, presenting a more attractive way to both teacher and student.

Finally, within the scope of the European partnership in the EasyICT project, the platform was also extended to allow each partner country to administer their own questions and students. This is an important feature request, as different countries may have slightly different questions and tests.

3.2 Test Questions

According to the test categories as described in Section 2, we identified three types of tests in the current pilot implementation:

1. Simulation test questions, including actions such as clicking and dragging. This test is performed in a simulation exercise (Flash).
2. 'In-application' tests include actions such as 'opening the browser', mostly performed by the candidate on their familiar desktop.



(a)



(b)

Fig. 1. (a)An example of a simulation question where the candidate has to drag the head of the cat into the right position. (b)Additional stimulating material for practicing moving the mouse, using the WAI-NOT website[6].

3. Finally, permanent evaluation tests are items that are observed during a longer observation period (e.g. appropriately handling the hardware).

Figure 1(a) shows an example of a simulation question, developed using Adobe Flash. Animations are made based upon drawings made by the children themselves. Animations and sounds are added in order to make it more attractive.

Finally, based upon the results of a complete test session, the candidates' profile is generated. In contrast to the report generated for the candidates themselves, this report is listing the acquired skills as well as the ones not yet acquired. Additionally, another report is generated containing a list of additional material that can be used for further development of the connected ICT skills. On the provided report, there is a list consisting of publicly available exercises that can be found on the internet, motivating the candidate and their tutor to start practicing on those particular skills (see figure 1(b)).

4 Conclusions and Future Work

In this paper we proposed the EasyICT project, an assessment framework and accompanying on-line assessment tool focussing on the assessment of the ICT skills of people with cognitive disabilities. As a result, the candidates get a positive report on their current skills, together with a personalised list in order to strengthen those skills. Currently a pilot test in five European countries is being performed.

Acknowledgements

EasyICT was co-funded by the Lifelong Learning Programme of the European Commission. The authors want to thank the EasyICT project partners: Academy of Management (Poland), Social Innovative Network Dr. Pretis (Austria), Kutahya Local Education Directorate (Turkey) and Applied Psychology Academy (Italy).

References

1. Website: EasyICT (2010), <http://www.easyict.eu>
2. Group, I.S.M.: E-Business and ICT skills in europe (2002)
3. Welsum, D.V., Vickery, G.: New perspectives on ICT skills and employment / working party on the information economy. Technical report, OECD, Paris (2005)
4. Website: ECDL foundation: The european computer driving licence (2010), <http://www.ecdl.com/publisher/index.jsp>
5. Website: Dokeos e-learning, general adress (2010), <http://www.dokeos.com/>
6. Website: WAI-NOT vzw (2010), <http://www.wai-not.org/>

Towards an Interactive Screening Program for Developmental Dyslexia: Eye Movement Analysis in Reading Arabic Texts

Areej Al-Wabil¹ and Maha Al-Sheaha²

¹ College of Computer and Information Sciences, Department of Information Technology

² College of Education, Department of Special Education

King Saud University

{aalwabil.c, alsheaha}@ksu.edu.sa

Abstract. In this paper, we describe how eyetracking has been used in exploratory experiments to inform the design of screening tests for dyslexic students by examining their eye gaze while reading Arabic texts. Findings reveal differences in the intensity of eye gaze and reading patterns between dyslexic readers and non-dyslexic controls. Dyslexics consistently exhibited longer fixation durations, shorter saccades, and more regressions. Moreover, results suggest that eye movement patterns are a reflection of the cognitive processes occurring during reading of texts in both Arabic deep and shallow orthographies. Applicability of eye movement analysis in investigating the nature of the reading problems and tailoring interventions to the particular needs of individuals with dyslexia is discussed.

Keywords: Dyslexia, Eye tracking, Arabic, Learning Difficulties, SpLD.

1 Introduction

Dyslexia is a Specific Learning Difficulty (SpLD) characterized by difficulties in reading, spelling, and writing. Research in dyslexia in Arabic has suggested that the unique characteristics of the language (bi-directional script with diacritics representing short vowels) contribute to reading difficulties in different ways [1]. Arabic is unique in terms of orthography and the use of diacritics in place of short vowels in shallow orthographies (e.g. texts used in learning materials and children's books), which are omitted in mainstream texts (e.g. newspapers, web sites) resulting in what is known as deep orthographies. Little is known about visual reading patterns in the Arabic language. Research suggests that cognitive processing in reading Arabic demands stronger visual spatial processing abilities [2]. The role of visual processing of Arabic texts is especially important to understand the nature of problems that individuals with developmental dyslexia experience in reading.

Eyetracking has been around for more than one hundred years. Research using eye tracking technologies can be categorized as interactive or diagnostic [3]. Reading was the key diagnostic research domain for eyetracking; it has contributed to our

understanding of how people read and how eyetracking measures are related to reading difficulties [4]. A number of eye tracking measures have been shown to be indicative of cognitive processes essential for reading. For example, backward saccades have been linked to weakness in short term memory, and intensity of gaze (e.g. how long an individual looks at a word) has been linked to the ease or difficulty associated with processing the target element. For example, longer fixations suggest increased difficulties in decoding and comprehending text by readers.

Over the past 30 years, abundant evidence has accumulated indicating that erratic eye movements of people with dyslexia are a reflection of a language processing deficit [4]. While recent advances have made eyetracking less obtrusive and accessible for research, most studies examined languages with Latin or non-alphabetic scripts and very few studies have applied it to examine reading Arabic scripts. Dyslexia is known to vary between individuals as it affects people in different ways and to varying degrees of severity; no two dyslexics are alike. The ability to screen dyslexic students by examining their reading patterns allows specialists to gain an insight to the type of problems that they have in order to address the remedial educational program accordingly.

In this paper, we describe how eyetracking was used to examine reading patterns of students with dyslexia and propose a method of using this to inform the individual learning support plans for specialists in SpLDs. Eye tracking can identify each student's reading skill strengths and weaknesses, so that specialists spend less time and effort on established skills and more time and support on areas of weakness, leading to fast, effective learning.

2 Method

An eye tracking experiment was conducted to explore the relationship between eye movement measures and reading patterns in Arabic scripts, and to inform the design of screening programs. For this study, eye movements were recorded using a Tobii x120 stand-alone eye tracker, with a sampling rate of 120 Hz and accuracy of 0.5 degrees. Freedom of head movement was 30x15x20cm at a viewing distance of 60cm.

2.1 Participants

Eight participants, 4 dyslexics and 4 non-dyslexic controls, took part in these exploratory experiments. Participants ranged in age between 10 to 12 years of age and were in grade levels ranging between grades 4 and 6.

2.2 Procedure

Participants were tested individually. Sessions started with conducting a 5-point calibration with the eye tracker. Following that, each participant completed one training task and four experimental tasks. The four tasks involved reading an Arabic passage at the third grade reading level, two of which were in deep orthography (no diacritics)

and two in shallow orthography (with diacritics). Fonts used were similar to the fonts used in children's reading books. Mappings from letters to sound are more consistent in the shallow orthography than the deep orthography. The reading was oral reading as opposed to silent reading so that the number and type of errors can be ascertained. Examples of a shallow orthography stimulus and a deep orthography stimulus with fixations overlaid are shown in Figure 1. The red circles represent eye fixations; the size of the circle is relative to the duration of the fixation. The red lines represent saccades, rapid eye movements between fixations. The direction of reading is indicated by the numbered sequence of fixations.

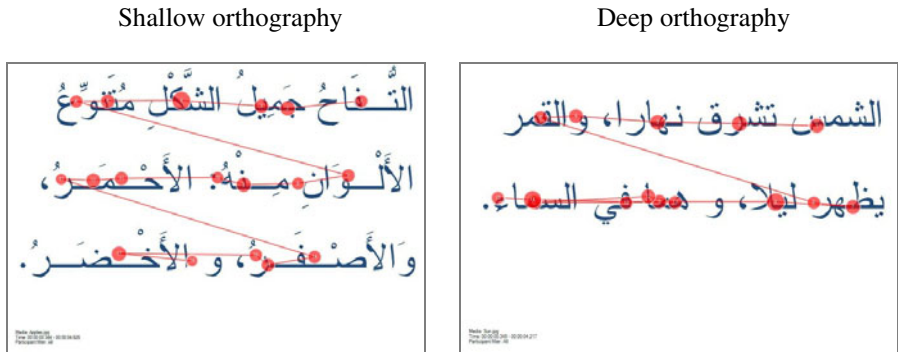


Fig. 1. Stimuli with reading patterns overlaid

Analysis of reading patterns is based on quantitative data of eye gaze allocation in different regions of the text. Regions include words, letters and diacritics for the stimuli with shallow orthography and words and letters for the stimuli with deep orthography. Reading errors were classified as either omissions or distortions. For omissions, the report discussed with SpLD specialists would describe whether the student fixated on the word/letter or not. For distortions and substitutions, we used the classifications of Savage Stewart and Hill [5] as well as (1) phonological (2) orthographic and (3) mixed and (4) unrelated category. Reading errors were further linked to descriptions of eye gaze such as intensity of gaze (e.g. did the reader exhibit intense fixations on the word suggesting difficulties in decoding text, did he/she not fixate on the word at all).

3 Results and Discussion

Results revealed marked differences in the reading patterns of students with dyslexia and controls in the context of Arabic text. An example demonstrating the variations in reading patterns between dyslexic readers and non-dyslexic readers is presented in Figures 2 and 3.

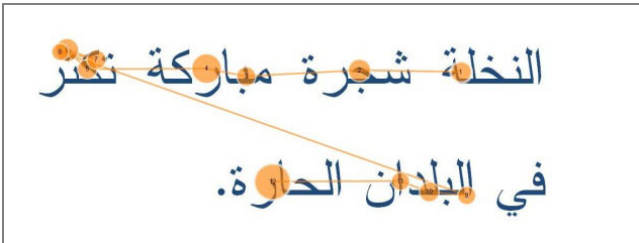
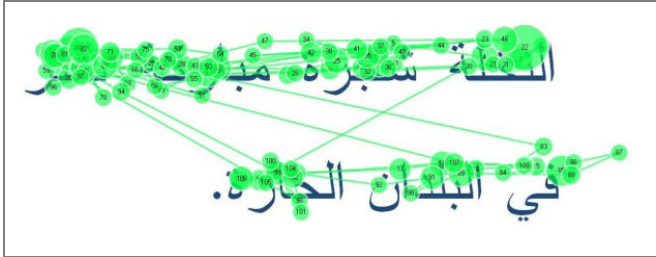
Reader	Deep Arabic Orthography (without diacritics)
Non-Dyslexic Control	
Dyslexic Reader	

Fig. 2. Visual patterns in reading deep orthographies by dyslexic and non-dyslexic readers

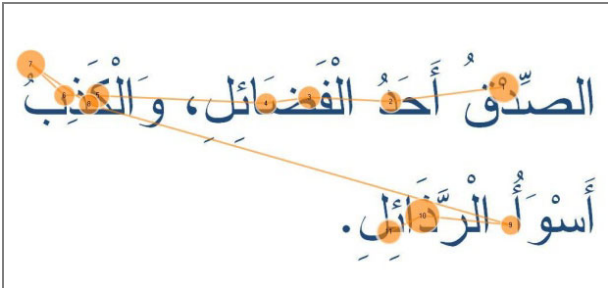
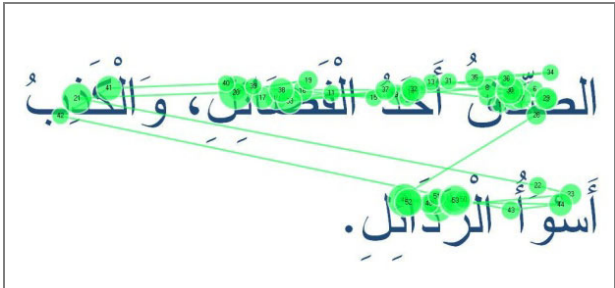
Reader	Shallow Arabic Orthography (with diacritics)
Non-Dyslexic Control	
Dyslexic Reader	

Fig. 3. Visual patterns in reading shallow orthographies by dyslexic and non-dyslexic readers

3.1 Eye Movements in Reading

Table 1 describes the eye movement measures exhibited by dyslexic readers and non-dyslexic readers. Each stimulus is listed in a column and the average is listed in the last column. Dyslexics consistently exhibited more fixations, longer mean fixations in reading and longer reading duration across all stimuli.

Table 1. Eye movement measures in reading Arabic texts by dyslexic and non-dyslexic readers

	Stimuli				Average
	1 (Tree)	2 (Apple)	3 (Sky)	4 (Truth)	
Reading time (in seconds)					
Dyslexics	16.46	12.80	10.10	19.04	14.60
Non-Dyslexic Controls	6.69	7.23	5.54	5.62	6.27
Number of Fixations (count)					
Dyslexics	49.25	39.75	33.50	65.50	47.00
Non-Dyslexic Controls	21.25	26.00	18.50	16.75	20.63
Mean Fixation Duration (in seconds)					
Dyslexics	0.346	0.310	0.322	0.330	0.327
Non-Dyslexic Controls	0.249	0.276	0.292	0.300	0.279

Students with dyslexia exhibited more fixations and longer fixations than controls suggesting increased difficulties in reading the text. Dyslexics also exhibited more backward saccades in their reading patterns than non-dyslexics. This was especially evident in the deep orthography in which readers need to understand the context of the word in the sentence in order to deduce the correct diacritics essential for properly pronouncing the Arabic words.

3.2 Reading Errors

Table 2 summarizes the number of errors that readers exhibited by participants in both groups in reading the passages. These include repetitions, additions, transpositions, omissions, substitutions. Examples of omissions and distortions are shown along with the eye gaze patterns in Table 3.

Table 2. Number of reading errors exhibited by participants (e.g. distortions, omissions)

	Stimuli				Total
	1 (Tree)	2 (Apple)	3 (Sky)	4 (Truth)	
Dyslexics	1	1	0	1	3
Non-Dyslexic Controls	7	9	7	4	27

Erratic patterns of eye movements were found in words that were distorted in pronunciation (e.g. out-of-sequence). In omissions, patterns of erratic eye movements preceding the element being omitted (e.g. letters or word) were exhibited. The second error illustrated in Table 3 shows how fixations were clustered within the boxed area in which the reader pronounced. It is noted that fixations were intense in this instance, thus sizes of circles were kept constant to indicate sequence in the image; however fixation duration in the boxed area suggests intense cognitive processing was exhibited by the reader for decoding that element.

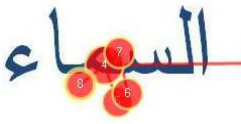




Dyslexics often add words in their reading that were not included in the script. It was interesting to note that additions of words by dyslexic participants were linked with patterns of a single relatively short fixation on the word before or after the inserted word, suggesting that the reader exhibit a sweep over the word and used context to guess what the words were at that point. These are illustrated in the fourth error described in Table 3.

While the notion of examining the fine details of eye movements in reading with children was a high-cost low-yield endeavor a few decades ago, recent advances in eye tracking technologies coupled with established theories of language processing that help interpret eye movement data have changed this. Findings of these experiments revealed the moment-to-moment cognitive processing activities of students in reading. This highlighted difficulties that they encounter for learning support specialists.

3.3 Profiles of Reading Problems

This data was presented to SpLD specialists in an accessible format such as video segments of the session and summary reports of erratic eye movement patterns, reading errors and possible interpretations to explore methods of profiling reading problems. This study's findings enabled specialists to obtain an insight into the child's abilities and cognitive profile and consequently tailor the intervention according to their reading difficulties. Insights obtained from this screening method can help the specialists determine the weaknesses and strengths of the student and plan a remedial program according to their profile.

Table 3. Examples of eye gaze patterns in errors exhibited in reading Arabic script

Error Description	Eye Gaze on Problematic Words	Word pronounced by dyslexic reader
1 Reversals were exhibited when eye gaze examined letters out of sequence. In this example the word pronounced by the reader matched the sequence of eye gaze.		المساء Reversal Masaa (evening) rather than Samaa (sky)
2 Omitting letters or words was linked to patterns of repeated backward saccades or re-examinations of areas prior to the element that was omitted.		البلد Letters beyond the green border were omitted
3 In some substitutions exhibited by readers, letters were examined out of sequence and with varying levels of intensity.		بلادنا Substitution
4 Additions were linked to patterns of scanning or quick sweeps across words.		في الليل Added 'fi'
		في النهار Added 'fi'

4 Conclusion

In order to develop effective intervention and remedial programs for students with reading difficulties, it is important to understand the problems that they experience in reading and identify the related cognitive difficulties. Recent advances in eye tracking have made it more accessible to obtain an insight at a high level of elaboration. The results of our comparative study between an experimental group and a control group show that reading patterns vary and are indicative of the type of problems that readers experience.

From a research perspective, findings suggest links between eye movement patterns and reading problems in Arabic. From a practical perspective, this study has demonstrated the invaluable assistance eye tracking can offer for SpLD practitioners in tailoring remedial programs for dyslexics according to their specific profile. This in turn can inform the design of screening programs for developmental dyslexia in Arabic. Ongoing work involves developing a coding scheme for developing a reading problems' profile for dyslexic students based on their eyetracking data. Future work would explore which texts are effective in revealing dyslexic difficulties, validating these with representative samples and developing a screening system for the various reading levels.

References

1. Abu-Rabia, S.: Reading Arabic Texts: Effects of Text type, Reader Type, and Vowelization. *Reading and Writing: An Interdisciplinary Journal* 10, 106–119 (1998)
2. Abu-Rabia, S.: Dyslexia in Arabic. In: Smyth, I., Everatt, J., Salter, R. (eds.) *International Book of Dyslexia: A Cross-Language Comparison and Practice Guide*, pp. 31–38. John Wiley & Sons, West Sussex (2004)
3. Duchowski, A.: *Eye Tracking Methodology: Theory and Practice*, 2nd edn. Springer, London (2007)
4. Rayner, K.: Eye Movements in Reading and Information Processing: 20 Years of Research. *Psychological Bulletin* 124, 372–422 (1998)
5. Savage, R., Stuart, M., Hill, V.: The role of scaffolding errors in reading development: Evidence from a longitudinal and a correlational study. *British Journal of Educational Psychology* 71, 1–13 (2001)

Developing a Multimedia Environment to Aid in Vocalization for People on the Autism Spectrum: A User-Centered Design Approach

Areej Al-Wabil, Hadeel Al-Shabanat, Rawan Al-Sarrani, and Manal Al-Khonin

College of Computer and Information Sciences, Department of Information Technology,
King Saud University
aalwabil.c@ksu.edu.sa

Abstract. This paper outlines the analysis and design stages of an interactive multimedia environment for encouraging vocalization of people with autism. We described the user-centered design approach for involving users in different roles in the design of an Arabic interactive rehabilitation tool adapted to the needs of speech therapy programs used in the local context. Needs analysis included exploratory surveys and interviews conducted to understand how technology can help in encouraging the vocalization of children with autism. The design stage involved iterative development with prototype evaluations with specialists to refine system functionality and ensure the accessibility and usability of the system. Insights from involving users in different roles are described.

Keywords: Autism, User-Centered Design, Vocalization, Arabic, Multimedia.

1 Introduction

Autism is a developmental disorder affecting social interaction and communication abilities. Interactive multimedia environments are increasingly being used to help individuals with such communication difficulties [e.g. 2]. Technology solutions have been shown to be effective in supporting people with autism either directly or in remedial sessions with specialists [3]. However, much of the programs and technology aids are designed with language and cultural considerations that vary considerably from the needs and requirements of native Arabic-speaking users and thus are inadequate for use in the local context which consequently results in low adoption rates. There is a lack of rehabilitation support for people with autism in Arabic-speaking populations. Accurate prevalence of Autism Spectrum Disorders (ASD) in many developing countries are limited, but it has been estimated at 18 per 10000 in Saudi Arabia [1]. Previous research has examined various approaches and solutions to encourage vocalization of people on the autism spectrum [e.g. 2,3]. The Mediate project involved developing a multimedia environment that aims to create a sensory dialogue with users by integrating three different modalities, visual, sound and tactile [2]. The spoken impact project involved designing an interactive program that responds to vocalizations of autistic children with graphic visualizations [3]. Our study extends investigatory work in the area, by probing deeper into social, cultural and language

design considerations for the local Arabic context, identifying new issues that need to be taken into account for the design of assistive technologies with Arabic interfaces and creating a novel solution to address these limitations.

2 System Design and Development

We conducted several studies to gain a deeper understanding of the communication challenges faced by people with autism and the professionals working to support them in the local context. Interactive systems that respond to vocalizations with multimedia visualizations and sounds (e.g. echo) were proposed and functionality was explored.

2.1 Needs Analysis

The needs analysis of target users, individuals with autism and specialists in rehabilitation programs were carried out through background surveys and interviews with parents of autistic children. Furthermore, usability and accessibility evaluations were conducted on a selected sample of software programs developed for autism to identify design considerations and understand the limitations of these programs for use in the local context that need to be addressed by our system.

An exploratory survey was conducted to understand how technology can help in encouraging the vocalization of children with Autism, and examine the types of interactive environments that attract and retain autistics' attention (e.g. sound, shape, animation, motion). Surveys were distributed to 55 specialists in local rehabilitation centers in Riyadh, which yielded 41 responses; a rather high response rate as specialists described how eager they were to participate to express their needs in terms of technology support. Furthermore, surveys were posted online in Arabic discussion forums for people with disabilities and support groups on social networks such as Facebook and views of parents and caregivers were elicited. Results indicated variability in preferences, reflecting individual differences in this target population which suggested that flexibility as a key feature in the system by enabling users to combine and adjust system settings of visualization, animation and sound settings. Findings also outlined the session data that is essential for specialists to keep track of progress for children with autism, such as level of engagement, level of enjoyment, types of utterances and quantitative measures of vocalizations.

2.2 Matching System Functionalities against Autistics' Needs Analysis

In this phase, we took the information collected in the needs analysis phase, which include the needs of children with autism, their caregivers and specialists supporting them, the survey questionnaire results, and our review of autism software and put that together in preparation for the design phase. This involved mapping key features to users benefiting from these features, together with a justification of why these features are relevant to specific users and whether they have been offered in similar systems developed for other contexts and the reported effectiveness if available (Table 1).

Table 1. Examples of mapping user's needs to system functionality

Characteristic / Requirement	Users	Functionality
Idiosyncratic preferences and habits	Autistic individual	Offering flexibility in the system design. Users can select the shape, animation path, backgrounds, and sound for adapting sessions according to the individual's preferences.
Evoking a sense of engagement	Rehab specialists, parents, caregivers	Accessible and usable interface for adjusting the settings of the environment and adapting the system's response to the preferences of the individual.
Sensitivity to sound	Autistic individual	Volume control and enabling users to select specific audio responses (e.g. echo, music)
Progress tracking and understanding an individual's profile of strengths and weaknesses	Rehab specialists	Qualitative measures are collected for each session such as levels of engagement and enjoyment, and types of utterances. Quantitative measures of vocalizations are mapped for analysis by speech therapists.

2.3 Design

In this phase, personas (types of users with common goals) were created to guide the development of the system, scenarios of use were formulated, and system functionality was determined. Functionality of the system was determined by using an exploratory approach using findings from our exploratory survey with specialists in rehabilitation centers, followed by a requirements discovery session conducted in a workshop with specialists. Prototyping was conducted to test discrete functionalities such as selecting graphics, adjusting sound settings, determining the type of animations in the multimedia environment.

2.3.1 Low-Fidelity Prototyping

Low-fidelity prototypes were examined in a hands-on session with speech therapists, education and remedial program specialists in a local autism support center. In these sessions, we explored how specialists in different domains envision the use of such systems and examined the specific features that they consider essential for their contexts of use. We also explored implementation alternatives such as handheld devices, full-immersion rooms, or large projection screens and elicited their views on the effectiveness of each alternative design.

2.3.2 High-Fidelity Prototyping

A high-fidelity prototype was demonstrated to target populations in a local symposium aimed at domain experts and rehabilitation specialists working with autistic individuals. Reviews were focused on interaction with the system, configurations, and scenarios of use in rehabilitation centers, particularly low functioning autistic children.

3 Conclusion

This paper describes the design and development of an interactive multimedia program to encourage vocalization of children with autism. System development is underway. The software will then be deployed with children with autism who receive speech therapy sessions to help them with their communication skills. The final evaluation phase will assess the efficacy of the multimedia environment in encouraging vocalization and communication in therapy programs.

Acknowledgments

Special thanks to all participants involved in our user research studies and to the staff of the Human Development Center and Prince Faisal's Center for Autism in Riyadh. We also thank project team members, Fatimah AlQahtani, Nojoom Al-Najem, Asma Al-Otaibi, and Noor Al-Othaim for their contributions throughout the project.

References

1. Al-Salehi, S., Al-Hifthy, E., Ghaziuddin, M.: Autism in Saudi Arabia: Presentation, Clinical Correlates and Comorbidity. *Transcult Psychiatry* 46(2), 340–347 (2009)
2. Gumtau, S., Newland, P., Creed, C., Kunath, S.: Mediate – A Responsive Environment Designed for Children with Autism. In: *Accessible Design in the Digital World Conference*, Dundee, Scotland (2005)
3. Hailpern, J., Karahalios, K., Halle, J.: Creating a Spoken Impact: Encouraging Vocalization through Audio Visual Feedback in Children with ASD. In: *Proceedings of the 27th International Conference on Human Factors in Computing Systems*, pp. 453–462. ACM, New York (2009)
4. Kientz, J., Hayes, G., Westeyn, T., Starner, T., Abowd, G.: Pervasive Computing and Autism: Assisting Caregivers of Children with Special Needs. *Pervasive computing* 6(1), 28–35 (2007)

The Performance of Mouse Proficiency for Adolescents with Intellectual Disabilities

Ting-Fang Wu¹, Ming-Chung Chen², and Chi-Fen Wu³

¹ Graduate Institute of Rehabilitation Counseling, National Taiwan Normal University
162, Sec. 1, Ho-ping E. Rd., Taipei, Taiwan, R. O. C.
tfwu@ntnu.edu.tw

² Department of Special Education, National Chiayi University
300 University Road, Chiayi, Taiwan, R. O. C.
mtchen@mail.ncyu.edu.tw

³ Department of Special Education, National Taiwan Normal University
162, Sec. 1, Ho-ping E. Rd., Taipei, Taiwan, R. O. C.
zoe9277@yahoo.com.tw

Abstract. Information and computer technology has grown rapidly and played an essential role in our education, vocation, and daily life. However, for students with intellectual disabilities, effective cursor control is challenged. The purpose of this study is to investigate the performance of mouse control of 10 adolescents with intellectual disabilities compared with their aged peer. A mouse proficiency assessment software was utilized to collect the data. The results indicated that the adolescents with intellectual disabilities who had mouse using experience do not perform as efficient as their peers without disabilities, although they could use the mouse with high accuracy rates. The adolescents with intellectual disabilities spend less reaction time, longer total time and movement time, larger ratio of PL/TA, more movement units to complete pointing and clicking tasks. The results provide essential reference for designers of computer assisted learning software when developing e-learning material for adolescents with intellectual disabilities.

Keywords: students with intellectual disabilities, pointing and clicking tasks, cursor control.

1 Introduction

Information and computer technology has grown rapidly and played an essential role in our education, vocation, and daily life. The skills of operating keyboards and pointing devices are prerequisite in successful academic learning and employment pursuing. With the popularization of graphical user interface, pointing devices and cursor control have become standard features of current computer system [1]. However, for students with intellectual disabilities, effective cursor control is challenged, particularly for those with severe impairments [2]. Li-Tsang, Yeung, and Hui-Chan found

only around one third of the 353 adults with intellectual disabilities could successfully use mouse to execute double clicking and dragging tasks (double clicking: 31.7%; dragging: 38.8%) [3]. Lin, Chen, Chang, Yeh, and Meng found pupils with intellectual disabilities performed worse than their aged matched peers [4]. The pupils with intellectual disabilities executed the tasks less accurately especially when the targets displayed in smaller size.

Studies in mouse performance have tended to focus on movement time and error rate as the principal interests. However, these 'gross measures' have limited abilities to describe the actual behaviors of the individual [1]. Accuracy and movement time are only the outcome indicators of mouse proficiency, rather than the causes of the performance difficulties. Mackenzie, Kauppinen, and Silfverberg proposed seven kinematics parameters as measurement indicators, including target re-entry (TRE), task axis crossing (TAC), movement direction change (MDC), orthogonal direction change (ODC), movement variability (MV), movement error (ME), movement offset (MO) [5]. Keates et al. [1] added two more parameters: missed click (MC) and ratio of path length to task axis length (PL/TA). Meng et al., selected reaction time, movement time, total path of trajectory, velocity, and movement unit(MU) as the parameters to measure[6].

In addition, Fitts' Law illustrated that size of the object and the distance of the object decide the difficulty of task, named as index of difficulty (ID) [7]. ID could be used to represent the size and distance of icons design on the user interface.

The purpose of this study is to investigate the performance of mouse control of adolescents with intellectual disabilities compared with their aged peer. The accuracy, total time, reaction time, movement time, adjusting time, rate of PL/TA, MU and target re-entry (TRE) were selected as the parameters to represent the mouse operating performance in 4 different levels of IDs.

2 Methods

2.1 Participants

Twenty adolescents aged from 15 years 5 months to 16 years 9 months old (with an average of 15 years 11 months old) were recruited from Neihu Vocational High School in Taipei, Taiwan. The participants of experiment group were ten students with intellectual disabilities who enrolled in the special education class. The participants of control group were recruited from the students of regular class at the same school with the matched sex and age of the experiment group. The participants of experiment group have to meet the following criteria: 1). with a diagnosis of mental retardation; 2). without hearing, visual or physical impairments; 3). able to follow oral instructions; 4). with computer experience with mouse; 5). able to maintain sufficient attention for at least 20 minutes. The IQ scores, measured by WISC- , of the experiment group are revealed in Table 1.

Table 1. The WISC- IQ scores of students with intellectual disabilities

Participant	Verbal IQ	Performance IQ	Full-scaled IQ
1	55	48	48
2	70	66	65
3	60	60	56
4	69	55	59
5	69	62	62
6	60	59	56
7	63	70	64
8	74	71	69
9	64	70	65
10	64	62	60
Mean	64.8	61.3	60.4
SD	5.71	7.32	6.02

2.2 Research Design

Two factors mixed design (2 x 4) were used to conduct the experiment. The between subjects factor was group, and the within subject factor was ID. There are four IDs generated from two distances (5 cm, 15cm) and two sizes of targets (1 cm, 1.5 cm). The types of the task were ID1= 2.737 (5cm, 1.5 cm), ID2= 3.322 (5cm, 1 cm), ID3= 4.322 (15cm, 1.5 cm), ID4= 4.907 (15cm, 1 cm). The accuracy, total time, reaction time, movement time, adjusting time, rate of PL/TA, movement units, and target re-entry (TRE) were selected as the parameters to represent the mouse operating performance for adolescents with and without intellectual disabilities. The definitions of these parameters are described as followings:

Accuracy refers to the percentage of correct responses. A correct response is defined as the participant moves the mouse cursor moving into the target area and clicks the left button right in the target area. The follow parameters were calculated from the correct responses only.

Total time is defined as the time the participants spent in completing a pointing and clicking task. It will be equal to the sum of reaction time, movement time and the adjusting time.

Reaction time (RT) is defined as the time from target display on the screen to the beginning of cursor movement.

Movement time (MT) is defined as the time from the beginning of cursor movement to the cursor reach the edge of a target.

Adjusting time (AT) is defined as the time it takes for the participants to adjust the mouse cursors to successfully complete the required tasks once the cursors moving into the target area.

Rate of PL/TA is defined as the actual length of the trajectory of the cursor movement divides the distance of the task. The higher rate indicates the less efficiency of the cursor movement.

Movement units (MU) occur at the phase the cursor accelerates and decelerates.

Numbers of target re-entry (TRE) represents the numbers of cursor re-entry the target before clicking. The more the numbers of movement units and target re-entry indicated the worse the mouse control.

2.3 Apparatus and Experimental Tasks

The pointing and clicking tasks were performed upon An ASUS PC with Celeron D346 processor 3.06 GHz equipped with 15 inches LCD monitor. A self-developed software, Computerized Assessment Tool (CAT), was used to measure and record the performance of mouse pointing and clicking [8, 9]. All participants were asked to use the same optical mouse to perform those pointing and clicking tasks. The tasks required the participants to move the mouse cursor to the target and click the left button when the cursor is in the target. The tasks of four different levels of difficulties (ID1, ID2, ID3, and ID4) displayed on the screen randomly. Each level of difficult includes 8 tasks of the same icon size and the same moving distance but in 8 directions (0° , 45° , 90° , 135° , 180° , 225° , 270° , and 315°). Therefore, each participant performed 32 pointing and clicking tasks for the experiment. CAT system recorded all the parameter automatically and ended when the 32 tasks were completed.

2.4 Experimental Procedure

The experiment was conducted individually in participants' school computer lab. Each participant was introduced the purpose of the evaluation and practiced to be familiar with both the procedure and the tasks. Then in a formal testing process, the participants were required to navigate the mouse cursor to the target and click as correctly and quickly as they could once the cursor was in the target.

2.5 Data Analysis

The means of the dependent variables, except accuracy, were calculated from the correct response represented the individual's performance. Two factors ANOVA with mixed design was used to analyze the data. The simple main effect of the groups and ID would be examined if the two factors interaction was significant. Otherwise, the main effect of groups and ID would be examined separately. The SPSS 13.0 was used to analyze the data. The multiple post hoc LSD tests were conducted to compare the difference between two IDs if the simple main effect or main effect examinations reach a statistical significance ($p < .01$).

3 Results

3.1 Accuracy

As the results disclosed in Table 2, the means of the accuracy for both group performing pointing and clicking tasks were over 96% across the four IDs, except that for students with intellectual disabilities performed the ID2. The interaction between group and task was not significant ($F=1.328$, $p=0.275$). Accuracy refers to the

percentage of correct responses. A correct response is defined as the participant moves the mouse cursor into the target area and clicks the left button right in the target area. The follow parameters were calculated only from the correct responses.

Table 2. The means and standard deviations of the performance in eight parameters for two groups

Variable	Group	ID ₁		ID ₂		ID ₃		ID ₄		Total Mean
		Mean	SD	Mean	SD	Mean	SD	Mean	SD	
Accuracy	1	1.000	0.000	0.913	0.167	0.976	0.051	0.964	0.058	0.963
	2	0.976	0.051	0.976	0.051	0.976	0.051	0.964	0.058	0.973
	total	0.988	0.037	0.9445	0.124	0.976	0.049	0.964	0.056	0.968
Total Time	1	1.387	0.128	1.397	0.126	1.625	0.242	1.567	0.114	1.494
	2	1.160	0.132	1.124	0.096	1.497	0.132	1.377	0.171	1.290
	total	1.274	0.172	1.261	0.177	1.561	0.201	1.472	0.172	1.392
RT	1	0.118	0.057	0.145	0.057	0.156	0.101	0.128	0.069	0.137
	2	0.243	0.099	0.225	0.038	0.271	0.150	0.247	0.078	0.247
	total	0.180	0.102	0.185	0.062	0.213	0.138	0.188	0.094	0.192
MT	1	0.654	0.091	0.733	0.120	0.871	0.197	0.879	0.156	0.784
	2	0.439	0.085	0.433	0.084	0.732	0.171	0.654	0.100	0.565
	total	0.546	0.139	0.583	0.184	0.802	0.193	0.766	0.172	0.674
CT	1	0.616	0.106	0.519	0.080	0.598	0.162	0.559	0.072	0.573
	2	0.478	0.124	0.466	0.076	0.494	0.055	0.476	0.131	0.479
	total	0.547	0.133	0.493	0.081	0.546	0.130	0.518	0.112	0.526
R_Path/Dis	1	2.334	0.960	2.169	0.476	1.409	0.140	1.457	0.345	1.842
	2	1.618	0.251	1.431	0.232	1.258	0.133	1.308	0.288	1.404
	total	1.976	0.776	1.800	0.526	1.333	0.154	1.382	0.319	1.623
MU	1	3.338	0.457	3.240	0.512	3.959	0.644	3.857	0.545	3.599
	2	2.321	0.417	2.166	0.383	2.957	0.639	2.779	0.922	2.556
	total	2.829	0.673	2.703	0.705	3.458	0.809	3.318	0.922	3.077
TRE	1	0.288	0.221	0.195	0.228	0.257	0.221	0.266	0.247	0.251
	2	0.127	0.157	0.158	0.125	0.175	0.186	0.160	0.265	0.155
	total	0.207	0.204	0.176	0.180	0.216	0.203	0.213	0.255	0.203

3.2 Total Time

The interaction effect was not significant ($F=1.177$, $p=0.327$). The results of simple main effect testing indicated that students with intellectual disabilities and students without intellectual disabilities performed differently in all four types of task ($F=20.338$, $p<0.001$) ($F=27.628$, $p<0.001$). The students with intellectual disabilities need longer time to complete all the tasks. For both groups, the results of post hoc LSD tests demonstrated that except the conditions of ID1-ID2 and ID3- ID4, all other four conditions were significantly different (Table 3). The participants, with intellectual disabilities or not, spent more time in completing the tasks when the distance was

longer no matter the sizes of targets. The results comply with the Fitts' law showed that a lawful relationship exists for the time taken to move to a target [7].

3.3 Reaction Time

The interaction effect was not significant ($F=0.475$, $p=0.701$). The main effect of group was significant ($F=14.719$, $p=0.001$). The main effect of ID was not significant ($F=0.031$, $p=0.862$). The students with intellectual disabilities spend less time to initiate their mice than matched peer across all 4 IDs tasks. That means students with intellectual disabilities may have poor impulse control, therefore, they initiate the actions as soon as the target displayed.

Table 3. The p value of the post hoc LSD tests

Variable	ID ₁ -ID ₂	ID ₁ -ID ₃	ID ₁ -ID ₄	ID ₂ -ID ₃	ID ₂ -ID ₄	ID ₃ -ID ₄
Total Time	0.636	0.000	0.000	0.000	0.000	0.065
MT	0.223	0.000	0.000	0.000	0.000	0.289
MU	0.259	0.000	0.006	0.000	0.001	0.508

3.4 Movement Time

The interaction effect was not significant ($F=2.483$, $p=0.071$). The results of simple main effect testing indicated that students with intellectual disabilities and students without intellectual disabilities performed differently ($F=22.177$, $p<0.001$) in all four types of task ($F=32.924$, $p<0.001$). Students with intellectual disabilities take longer movement times to complete the tasks in all four conditions. The results of post hoc LSD tests disclosed that participants, with intellectual disabilities or not, spent more times in completing the tasks when the distances were longer. The results were similar with the findings of Lin and his colleague (2009).

3.5 Adjusting Time

The interaction effect was not significant ($F=0.628$, $p=0.600$). The main effect of group was significant ($F=12.488$, $p=0.002$). The main effect of ID was not significant ($F=1.327$, $p=0.274$). The results indicated that students with intellectual disabilities need longer time to adjust the mouse cursors when the cursors reaching the target areas.

3.6 Rate of PL/TA

The interaction effect was not significant ($F=3.399$, $p=0.024$). The main effect of group was significant ($F=14.719$, $p=0.001$). The students with intellectual disabilities performed worse than matched peer on the mean. The main effect of ID was not significant ($F=5.950$, $p=0.025$). The students with intellectual disabilities were not as efficient as their aged peer in controlling mouse cursors by indicating the larger rates of PL/TA, especially for those short distance tasks.

3.7 Movement Units

The numbers of movement units represent the degrees of smoothness of cursor movements. The interaction effect was not significant ($F=0.030$, $p=0.993$). The results of simple main effect tests showed that both groups ($F=36.116$, $p<0.001$) performed differently among the four types of task ($F=10.935$, $p<0.001$). The students with intellectual disabilities were not as efficient as their aged peer in controlling mouse cursors by indicating the more movement units in completing tasks.

3.8 Numbers of TRE

The interaction effect was not significant ($F=0.428$, $p=0.734$). The main effects of group and ID were also not significant ($F=2.188$, $p=0.156$) ($F=0.220$, $p=0.882$). Lin et al. [4] when the tasks were the same ID, the students with intellectual disabilities had more overshoots in shorter distance but smaller target [4]. Furthermore, when the targets were bigger, they had more TRE in longer distance; however, the performance indicated no significant difference when the targets were small.

4 Conclusions

The results of this study indicated that the adolescents with intellectual disabilities who had mouse using experience do not perform as efficient as their peers without disabilities, although they could use the mouse with high accuracy rates. The adolescents with intellectual disabilities spend less reaction time, longer total time and movement time, larger ratio of PL/TA, more movement units to complete pointing and clicking tasks.

Lin et al. [4] found that size of the target plays a significant role in affecting the performance of mouse pointing and selecting for the students with intellectual disabilities. However, we did not reach the same trend in current study. The participants from Lin's study were 5th and 6th graders from elementary school; and the participants in current study are high school students. The effect of target size may diminish since the age increase. Further studies may recruit the students with intellectual disabilities with different ages and replicate the research with different levels of ID. The results provide essential reference for designers of computer assisted learning software when developing e-learning material for adolescents with intellectual disabilities.

References

1. Keates, S., Hwang, F., Langdon, P., Clarkson, P.J., Robinson, P.: The use of cursor measures for motion-impaired computer users. *Universal Access in the Information Society* 2, 18–29 (2002)
2. Wehmeyer, M.L.: Assistive technology and students with mental retardation. *Journal of Special Education Technology* 14(1), 48–58 (1999)
3. Li-Tsang, C.W.P., Yeung, S.S., Hui-Chan, C.: Factors affecting people with intellectual disabilities to learn computer technology. *International Journal of Rehabilitation Research* 28, 127–133 (2005)

4. Lin, Y.L., Chen, M.C., Chang, U.T., Yeh, C.C., Meng, L.F.: The performance of mouse pointing and selecting for pupils with and without intellectual disabilities. *Research in Developmental Disabilities* 30, 1188–1195 (2009)
5. Mackenzie, I.S., Kauppinen, T., Silfverberg, M.: Accuracy measures for evaluating computer pointing devices. In: *Proceedings of the 2001 Conference on Human Factors in Computing Systems*, pp. 9–15. ACM Press, Seattle (2001)
6. Meng, L.F., Chen, M.C., Chu, C.N., Wu, T.F., Lu, C.P., Yeh, C.C., Yang, C., Lo, C.Y.: The directional effects on cursor dragging kinematics. *International Journal of Computer Science and Network Security* 7, 1–6 (2007)
7. Fitts, P.M.: The information capacity of the human motor system in controlling the amplitude of movement. *Journal of Experimental Psychology* 47, 381–391 (1954)
8. Chen, M.C., Wu, T.F., Chu, C.N.: Essential competencies of special educators in integrating assistive technology into curriculum. In: Miesenberger, K., Klaus, J., Zagler, W.L., Burger, D. (eds.) *ICCHP 2004. LNCS*, vol. 3118, pp. 171–177. Springer, Heidelberg (2004)
9. Chen, M.C., Chu, C.N., Wu, T.F., Yeh, C.C.: Computerized assessment approach for evaluating computer interaction performance. In: Miesenberger, K., Klaus, J., Zagler, W.L., Karshmer, A.I. (eds.) *ICCHP 2006. LNCS*, vol. 4061, pp. 450–456. Springer, Heidelberg (2006)

When Words Fall Short: Helping People with Aphasia to Express

Tom Koppenol, Abdullah Al Mahmud, and Jean-Bernard Martens

Department of Industrial Design, Eindhoven University of Technology
5600 MB Eindhoven, The Netherlands
{a.al-mahmud, j.b.o.s.martens}@tue.nl

Abstract. In this paper we present the design of an application that helps people with expressive aphasia to express their feelings and share information through digital photographs. Our design is based on feedback from the therapist and persons with aphasia and their partners. The preliminary prototypes are evaluated with persons with aphasia and their partners. The concept is well accepted by the aphasics and it could be easily used during therapy and post-therapy period.

Keywords: Aphasia, design, storytelling, digital photos, communication.

1 Introduction

As a human being we communicate with others and share several things such as past events, feelings etc. Communicating verbally helps us to express our intention and makes us understandable to others. However, not all people in our society are able to express verbally such as people suffering from expressive aphasia. Aphasia is a language disorder, causing people to become impaired to express through reading, writing or speaking [5]. Augmentative and Alternative communication (AAC) devices have been built to help people with aphasia to communicate basic needs and wants. AAC devices mostly utilize icons for communication. However, icons are limited to express wide range of communication needs. Next to AACs new forms of communication support systems are emerging for people with aphasia [1].

Digital pictures or images contain much more information which are also a good source of communication support if it can be utilized properly [2, 3]. Digital photos have been utilized for several other purposes such as telling stories, as a cognitive assistant [4, 6, 8]. However, most of these applications are not tested for aphasics who have unique problems in reading, writing and or speaking. Moreover, how to utilize digital photos to express wide range of communication needs such as ‘needs and wants’ and also to ‘share information’ by utilizing computing technology have not been explored particularly for aphasics. Therefore, we propose to design an application by utilizing digital photos that serves two purposes: a) helps people with aphasia for needs-based interaction and b) helps them to share information such as short stories. The designed application adds additional support such as a) categorized menus based on ‘who’, ‘where’, ‘when’ and ‘what’ that are suitable for aphasics, and b) use of digital pictures to create a context and convey a significant meaning to communication partners.

2 Concept Design

We interviewed one speech therapist to understand how aphasics are diagnosed, how and what communication aids they use. Based on the interview results a concept called ‘Virtual World’ is created. We would like to enable aphasics to quickly create their own pictures as unrestricted as possible to show to others. We used emoticons as emotion play a facilitatory role in the production of communication in language-impaired people [7]. Two variations of the ‘Virtual World’ concept were created.

Interface 1. The interface consists of one main page (Fig. 1-left). There are 4 buttons to open submenus for ‘Who’, ‘What’, ‘Where’ and ‘When’. On the right side there is a bar to depict the emotion of the aphasic. All submenus are shown in blue over the picture. After clicking on an image, the image appears on the screen. In the ‘Where’ menu the aphasic can choose the background of the scene. Through the ‘Who’ and ‘What’ the aphasic can select persons and objects which are important for the picture. These can be dragged across the screen. The time is given in words and can be changed through the ‘When’ menu. The aphasic can alter the ‘emotion’-bar to approve a certain context.

Interface 2. The menu of interface 2 (fig. 1-right) is placed on the left side of the screen and consists of two layers. The menu buttons consist of pictures and images representing the submenu. When for instance the emotion menu is chosen, the ‘main menu’ panel will be replaced by the ‘emotion menu’ panel. Instead of the abstract emotion bar in Interface 1, the aphasic can import or delete ‘emoticons’. When the aphasic wants to show the time he can click on the ‘Time’ button (the clock). The time is indicated at the bottom of the screen and can be changed by pressing the left arrow for earlier and the right arrow for later. The user may ask questions by clicking on the question mark. A couple of questions appear on the screen. Thus the aphasic can select a question to show to his conversation partner.

In both concepts, the digital device virtually depicts the user and an environment with digital pictures. The user is able to display what he has experienced or what he wants or is going to do. The user can import and drag real pictures of objects and people.



Fig. 1. Interface 1 (left), Interface 2 (right)

Feedback and redesign. Both versions were tested with the speech therapist. The therapist appreciated the concepts as aphasics could be in control of the entire context. Based on the therapist's feedback the interface is redesigned (Fig. 2). New features in the interface are the 'story-telling' and the 'needs' tool. In the story-menu the user can select stored pictures and place them in the preferred order. In the needs-menu, commonly used or preferred needs are depicted. These can be direct questions or words. The final interface is designed with Adobe Flash and ActionScript.

The interface was evaluated with a 65-year old man who had a mild form of expressive aphasia. According to him the program covered a sufficient amount of features. The picture menu might be too complicated for some people with a severe form of expressive aphasia. However, when the program is understood, it would be very useful. Dragging people and objects is useful and can be learned. He mentioned that aphasics would use the product at home with family just like photo-albums to show their stories and communicate their experiences.



Fig. 2. Final prototype: main menu (left), need tool (middle), the story tool (right)

3 User Testing

Based on the feedback, two more refined final prototypes were evaluated with two aphasics (P1 and P2) who were recruited through a local rehabilitation center. P1 had a severe form of aphasia and P2 had a mild form of aphasia. First, we did an assessment and collected background information about the participants. Later on, both interfaces were demonstrated to the participants. Accordingly they were asked to provide feedback such as likes and dislikes through yes/no questions. Family members were also present during the feedback sessions and they were also asked to comment on the interface.

From P1, we learned that both interfaces were very clear for him to understand. Through the questionnaire he let us know that he could create a story and he could express through the program. He would also be able to express his emotion and his opinion. Moreover, he was sure that he would be able to ask relevant questions through the interface. However, he thought that the 'picture' menu had too many options to choose and menu buttons were too small. P2 mentioned that the product was well designed and aphasics would be able to tell each other about their experiences and they could communicate their needs. He appreciated the interface since it might guide the aphasic in creating a picture. He thought that it would require some time to learn the interface. However, aphasics would surely benefit from altering the pictures by dragging and changing the captions.

4 Discussion and Conclusion

Though two of the users would not require any communication device still they mentioned that people with expressive aphasia would benefit from this application. It provides sufficient features in a way that the interface is not difficult to use. The speech therapist also appreciated the concept. P2 mentioned that the prototype used a combination of words and pictures and so it had the ability to help aphasics to support their speech. The user with severe expressive aphasia (P1) thought that the interfaces were not complex in use. This might be the down side of demonstrating a product. Therefore future user tests should also cover the aspect of ‘trying out’.

From the early design phase proxy users evaluated the concept which was beneficial. The design process was user-centered and based on iterative concept development. The prototype used a combination of words and pictures and so it had the ability to help aphasics to express wide range of needs. People with severe expressive aphasia may benefit from the product’s innovative way of creating pictures and communicating needs, emotions and experiences. As a next step we intend to formally evaluate the prototype with severe expressive aphasics.

References

1. Al Mahmud, A., Martens, J.B.: Re-connect: Designing Accessible Email Communication Support for Persons with Aphasia. In: CHI EA 2010, pp. 3505–3510. ACM Press, New York (2010)
2. Al Mahmud, A., Limpens, Y., Martens, J.B.: Expressing Through Digital Photographs: An assistive Tool for Persons with Aphasia. In: 5th Cambridge Workshop on Universal Access (UA) and Assistive Technology (AT), CWUAAT 2010, pp. 157–166. Springer, London (2010)
3. Allen, M., McGrenere, J., Purves, B.: The Field Evaluation of a Mobile Digital Image Communication Application Designed for People with Aphasia. *ACM Trans. Access. Comput.* 1(1), 1–26 (2008)
4. Danielsson, H., Svensk, A.: Digital Pictures as Cognitive Assistance. Assistive technology: added value to the quality of life. In: AAATE 2001, pp. 148–152 (2001)
5. Hillis, A.E.: Aphasia: Progress in the last quarter of a century. *Neurology* 69, 200–213 (2007)
6. Jönsson, B., Svensk, A.: Isaac-A Personal Digital Assistant for the Differently Abled. In: 2nd TIDE Congress, pp. 356–361. IOS Press Inc., Amsterdam (1995)
7. Lorch, M.P., Borod, J.C., et al.: The role of emotion in the linguistic and pragmatic aspects of aphasic performance. *Journal of Neurolinguistics* 11(1-2), 103–118 (1988)
8. Mckelvey, M., Dietz, A., Hux, K., Weissling, K., Beukelman, D.: Performance of a Person with Chronic Aphasia using Personal and Contextual Pictures in a Visual Scene Display Prototype. *Journal of Medical Speech Language Pathology* 15, 305–317 (2007)

MarkerMouse: Mouse Cursor Control Using a Head-Mounted Marker

Rados Javanovic and Ian Scott MacKenzie

Dept. of Computer Science and Engineering
York Univeristy
Toronto Ontario Canada M3J 1P3
{rados,mack}@cse.yorku.ca

Abstract. We propose *MarkerMouse*, an inexpensive method for controlling the mouse cursor using a web cam and a marker placed on the user's forehead. Two modes of cursor control were compared: position-control and velocity-control. In position-control mode the cursor is positioned where the user's head is pointing. In velocity-control mode the mouse cursor moves in a constant speed in the direction the user's head is pointing. In an experiment designed according to ISO 9241-9, we found a mean throughput 1.61 bps in position-control mode. Throughput was 34% less, or 1.07 bps, in velocity-control mode. We explain how from the marker image we control the mouse cursor position and reduce noise in our computations.

Keywords: User interfaces, cursor control, web cam, marker tracking, head position tracking, head-operated mouse, mouse emulation, ISO 9241-9.

1 Introduction

Interacting with desktop computers typically engages both hands of the user. This is not a problem for most people, but for people with certain disabilities this scenario may be untenable. Fortunately, there is considerable research on providing efficient access to computers for disabled users. Several interesting technologies and solutions have emerged such as blink-based text entry [7] and voice-based mouse cursor control [4].

People with severe disabilities who cannot use, or have limited use of, their hands and who cannot speak might not be able to control an on-screen cursor. Sometimes the only option is to use the motion of the head to control the cursor. However, commercial solutions are typically expensive. Origin Instruments' *HeadMouse Extreme*, for example, costs about one thousand U.S. dollars (<http://www.orin.com>). A solution that costs less would be highly desirable. A new method for head cursor control is proposed in this paper.

Our proposed method does not require expensive hardware. It uses a common web camera, a printout of a marker (Fig. 1a) mounted on the user's forehead, and the software to process the marker orientation and move the cursor. Notable, as well, is that the user-to-computer interface is wireless, making it comfortable to use. We refer to this head-mounted marker system for mouse cursor control as *MarkerMouse*.

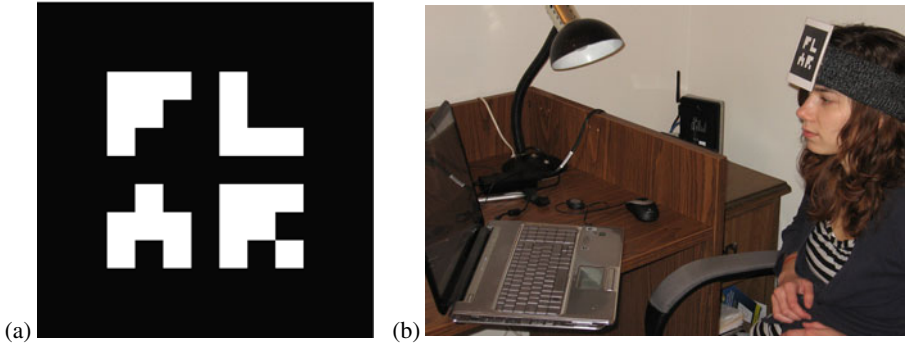


Fig. 1. *MarkerMouse* (a) printout (b) participant using *MarkerMouse*

The system is intended for people with restricted motor-control in the arms since only head motion is required to move the mouse cursor. The rest of this paper is organized as follows. In the next section, we review related work. Following this, we describe our system and provide details on the interface and noise reduction algorithm. Finally, we present an evaluation of our system, designed according to the ISO 9241-9 standard for evaluating non-keyboard input devices [5].

2 Related Work

Most uses of head-mounted markers are in virtual worlds to connect the physical position and movements of a real person with the virtual objects they interact with. For example, Knoeflein et al. [6] used a head-mounted marker in a virtual world gaming application. An IR marker was attached to a head-mounted camera and tracked to determine the user's head pose while playing ping-pong. Applications for cursor control tend to use face and feature tracking, rather than markers [2, 11]. In these applications, the focus is on the algorithms rather than the interaction. The use of a head-mounted marker greatly simplifies the software, since the geometric features in the marker are both precise and user-independent.

3 MarkerMouse

In this section, we describe the *MarkerMouse* interface, the position-control and velocity-control modes of cursor control, button clicks, and noise reduction.

3.1 Interface Details

The *MarkerMouse* head-mounted marker system is simple to assemble. The user simply prints the marker, positions it in some manner on her forehead, installs the software, and performs a simple calibration procedure. The marker can be attached to the user's glasses, a hat, a bandana, or anything that might be comfortable.

For *MarkerMouse* to work, the user must sit in front of a web camera so that the marker on the user's forehead is clearly visible (Fig. 1b). The orientation of the marker with respect to the camera when the user is looking straight at the monitor is the "neutral orientation". The neutral orientation is sampled and used in computing the movement of the marker and, hence, the motion of the mouse cursor.

3.2 Image Processing

The software linking the marker orientation to mouse cursor movement processes the marker image as input by the web camera. The first task is to compute the marker's normal vector in 3D space. Since the marker has a distinct, non-symmetric pattern on a square background it is possible to take an image captured by the web camera and retrieve the 3D plane on which the marker is sitting. By finding this plane, we calculate the normal vector of the marker. This is the neutral orientation. When the marker undergoes movement, we take the angle between the new normal vector and the neutral orientation and use this to determine where to move the cursor. To compute the marker's normal vector we used a software library called *NyARToolkit*¹.

3.3 Button Clicks

Our primary focus is on emulating mouse cursor control, apart from the functionality of clicking. To emulate a mouse button click, our system uses any key press on the system's keyboard. For accessible computing, depending on the disability, users may have enough manual facility to press a switch, for example using their hand. For users with severe disabilities who are not able to use their arms, the mouse button clicks can be emulated in different ways, for example, using blinks [7, 10], intentional muscle contractions [1, 3] or a sip-and-puff straw [12].

Position-Control and Velocity-Control

We evaluated two different modes of cursor control: position-control and velocity-control. In position-control mode, the marker orientation on the user's head determines the final position of the cursor on the screen. For example, if the user looks straight at the monitor the cursor will be in the center of the screen. If the user looks left the cursor will be on the left side of the screen, and if he then looks straight ahead again the cursor will be in the center again. This is similar to the operation of a stylus on a tablet or a finger on a touchscreen.

In velocity-control mode, the head orientation determines the direction to move the cursor. If the user looks straight at the monitor, the cursor is motionless. If the user looks left the cursor starts moving left and continues moving left while the user continues to look left. Movement is initiated when the computed displacement is equivalent to 200 pixels of mouse displacement. The rate of cursor movement is constant at 150 pixels per second. On the test system, this amounts to 10% horizontally and 16% vertically of the display extent per second. In velocity-control mode, we constrained the mouse movement to eight directions: up, down, left, right, and along the four diagonals.

¹ Available from Virtual Worldlets (<http://www.virtualworldlets.net/>).

3.4 Noise Reduction

Having the mouse cursor position directly dependent on the orientation of the marker may cause the cursor to jump around due to camera noise and inaccuracies in computing the marker's transformation. We used two methods to minimize noise.

The first method uses dampening. Instead of using the marker's normal vector to position the cursor directly, we compute the normal vector as a tentative cursor position. For each step the cursor moves accordingly, but with added dampening, as follows. Before the cursor is moved, a vector is computed in the direction of movement. The length of the vector is the distance between the two positions. This vector is multiplied by a dampening constant, D ($0 < D < 1$), and then the cursor takes a step accordingly:

$$NewMousePos = MousePos + (TargetPos - MousePos) \times D \quad (1)$$

In our application, we set D to 0.5. For a large head movement, the cursor moves relatively fast. For a small head movement, the cursor moves somewhat more slowly. This helps reduce jitter in the cursor movement due to small variations in the computation of the marker's normal vector.

The second method uses a history of the mouse cursor position. Previous cursor positions are recorded and used in computing the new position. Each new cursor position was determined using the computed, new position averaged with the last 10 mouse positions. This sliding-average approach helps minimize big spikes in the change of the marker's normal vector due to noise.

Applying these two methods together, the mouse cursor behaviour is more stable.

4 Method

While the design and initial testing of *MarkerMouse* system were promising, a proper evaluation with a pool of participants is required to validate the design. Our experimental evaluation tested how well participants could execute point-select tasks with our *MarkerMouse* in each mode of cursor control. The evaluation also included a generic mouse as a baseline condition.

Participants. Nine able-bodied participants were recruited from the local university campus. Five were male, four female, Ages ranged from 22 to 27 years. All participants used a computer on a regular basis.

Apparatus (Hardware and Software). Participants worked on a Hewlett Packard *Pavilion dv7* laptop computer. The system includes an embedded web camera on the top of the display. The resolution of the display was 1440×900 pixels. The operating system was Microsoft *Vista* (Home Edition).

Two separate software utilities were used. For demonstrating *MarkerMouse* and to provide participants the opportunity to practice, a simple program presenting an array of circular targets was created. See Fig. 2a. The targets were of random diameters in the range 30-80 pixels and were randomly scattered about the display. The targets could be selected in any order. When a target was clicked on, the colour changed to green to give visual feedback of a successful selection.

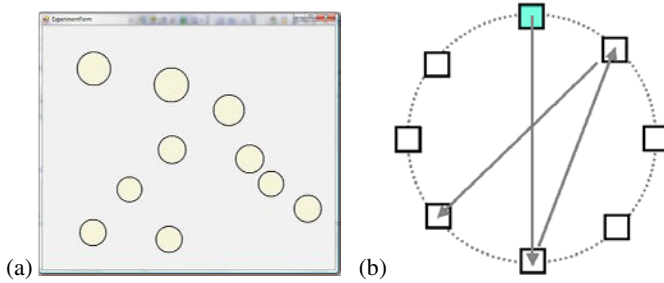


Fig. 2. Software (a) practice (b) ISO 9241-9 task #2

The experiment software was an implementation of task #2 in ISO 9241-9. See Fig. 2b. Eight square targets were uniformly positioned around a layout circle with a diameter of 600 pixels. Each target was 80×80 pixels. The participant first clicked on the top target, then on the target directly opposite, then on the target next to the first target, and so on around the layout circle. At any time, the target to select was highlighted. The target colour changed during selection. Selecting all targets was a “sequence”. This was repeated five times per mode of control.

Procedure. Participants sat in front of the laptop computer. The marker was put on the participant’s forehead clearly visible to the web-cam. The participant was asked to look straight into the screen and with a press of a button the current orientation of the marker was saved. This orientation was the neutral orientation and was used for system calibration. The calibration was done only once before each condition. Before each condition, participants spent 10 minutes getting familiar with the head-mounted marker using the practice software. For each condition, the participant was instructed on how the mode of control worked and on what was expected. Testing started after any button was pressed.

After the completion of all tests a questionnaire was handed to the participants to solicit their impressions on the input modes. The questionnaire was based on response items in ISO 9241-9 [5]. Each item had a rating from 1 to 5, as follows:

- Smoothness during operation (1: very smooth – 5: very rough)
- Mental effort required for operation 1: too low – 5: too high)
- Accurate pointing (1: easy – 5: difficult)
- Neck fatigue (1: none – 5: very high)
- General comfort (1: very uncomfortable – 5: very comfortable)
- Overall the input device was (1: very difficult to use – 5: very easy to use)

Experiment design. The experiment was a 3×5 within-subjects design, with the following independent variables and levels:

- Mode of control: Position-control, Velocity-control, Generic Mouse
- Sequence: 1, 2, 3, 4, 5

Participants were divided into three groups with the mode of control conditions administered using a Latin square.

The experiment included throughput as the dependent variable, as specified in ISO 9241-9. Throughput, in bits per second (bps), is a composite measure combining the speed and accuracy of participant responses. After each sequence of trials, throughput was calculated using

$$\text{Throughput} = ID_e / MT \quad (2)$$

MT is the mean movement time per trial over all trials in a sequence. ID_e is the effective index of difficulty, computed as

$$ID_e = \log_2(D / W_e + 1) \quad (3)$$

where D is the distance in pixels between the two targets (600 pixels) and W_e is the width of the distribution of clicks. W_e , is computed using

$$W_e = 4.133 \times SD_x \quad (4)$$

where SD_x is the standard deviation in selection coordinates along the axis of approach. Details are provided in other papers [e.g., 9].

5 Results and Discussion

Throughput. As clearly seen in Fig. 3a, there was a significant effect of mode of control on throughput ($F_{3,8} = 555.7$, $p < .0001$). Not surprisingly, the mouse performed best. Its throughput of 4.42 bps is consistent with mouse results in other ISO 9241-9 evaluations [9], thus validating our methodology.

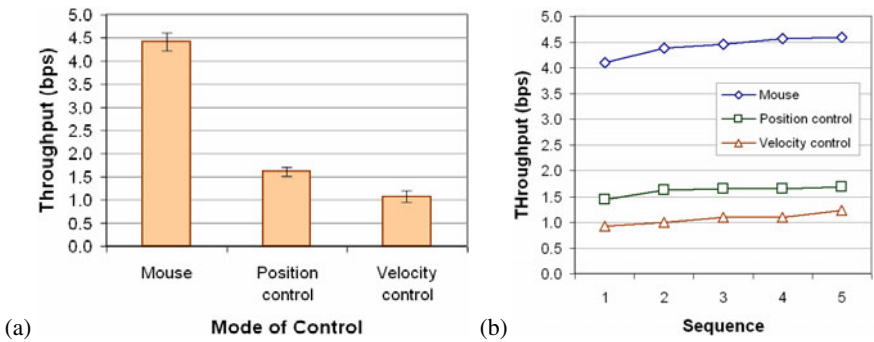


Fig. 3. Throughput (a) by mode of control (b) by sequence

The mean throughput for position-control movement of 1.61 bps was about 34% higher than the throughput of 1.07 bps observed for velocity-control movement. While substantially less than the throughput for the mouse, these figures are consistent with throughput for other pointing devices, such as a joystick (1.8 bps; see [8]).

As the trials continued, the throughput for both the position-control and velocity-control modes improved (Fig. 3b). After the first sequence, the position-control mode did not improve much, but the throughput for the velocity-control mode kept increasing.

Position-control mode produced better results because the mouse cursor went to the target position determined by the marker orientation almost instantaneously. In velocity-control mode, the mouse cursor had a constant speed (150 pixels/second). This slowed velocity-control mode, as multiple iterations of move and adjust were needed to position the cursor on the target.

Participant feedback. As *MarkerMouse* is intended for accessible computing and testing was done with able-bodied participants, there was an expected bias favoring the mouse. In comparing the two modes of control for *MarkerMouse*, participants indicated an overall preference for the position-control mode, rating it better than velocity-control mode for both neck fatigue and accurate pointing. Participants, however, gave slightly better scores to velocity-control mode for mental effort and smoothness.

6 Conclusions

In this paper we introduced a new method of controlling a mouse cursor using a marker placed on the user's forehead. Two methods of cursor control, position-control and velocity-control, were compared along with a desktop mouse. The mean throughputs for the position- and velocity-control modes were 1.61 bps and 1.07 bps, respectively, compared to the desktop mouse mean throughput of 4.42 bps.

Noise in marker detection was a limitation for the position-control mode. Velocity-control mode was less influenced by noise. However, because velocity-control is less intuitive, velocity-control mode performed worse than position-control mode.

An improvement for position-control movement would be a better way to deal with the marker detection noise, to improve precision and reduce jitter. This problem can be approached from two directions: to have a more robust method for marker detection and marker orientation calculation, and to have a better smoothing algorithm that minimizes noise after the orientation was computed.

References

1. Barreto, A.B., Scargle, S.D., Adjouadi, M.: A practical EMG-based human-computer interface for users with motor disabilities. *Journal of Rehabilitation Research* 37, 53 (2000)
2. Chauhan, V., Morris, T.: Face and feature tracking for cursor control. In: *Proceedings of the Scandinavian Conference on Image Analysis* (Unpublished, 2001); retrieved from <http://www.cs.man.ac.uk/~tmorris/SCIA630.pdf> (December 23, 2009)
3. Felzer, T., Nordmann, R., Rinderknecht, S.: Scanning-based human-computer interaction using intentional muscle contractions. In: Smith, M.J., Salvendy, G. (eds.) *Universal Access in Human-Computer Interaction, Part II, HCII 2009*. LNCS, vol. 5618, pp. 309–318. Springer, Heidelberg (2009)
4. Harada, S., Landay, J.A., Malkin, J., Li, X., Bilmes, J.A.: The vocal joystick: Evaluation of voice-based cursor control techniques. In: *Proceedings of the ACM Conference on Computers and Accessibility - ACCESS 2006*, pp. 187–204. ACM, New York (2006)
5. ISO, Ergonomic requirements for office work with visual display terminals (VDTs) - Part 9: Requirements for non-keyboard input devices (ISO 9241-9), International Organisation for Standardisation. Report Number ISO/TC 159/SC4/WG3 N147, February 15 (2000)

6. Knoerlein, B., Szekely, G., Harders, M.: Visuo-haptic collaborative augmented reality ping-pong. In: Proceedings of the International Conference on Advances in Computer Entertainment Technology - ACE 2007, pp. 91–94. ACM, New York (2007)
7. MacKenzie, I.S., Ashtiani, B.: BlinkWrite: Efficient text entry using eye blinks. In: Universal Access in the Information Society (UAIS) (in press)
8. MacKenzie, I.S., Kauppinen, T., Silfverberg, M.: Accuracy measures for evaluating computer pointing devices. In: Proceedings of the ACM Conference on Human Factors in Computing Systems - CHI 2001, pp. 119–126. ACM, New York (2001)
9. Soukoreff, R.W., MacKenzie, I.S.: Towards a standard for pointing device evaluation: Perspectives on 27 years of Fitts' law research in HCI. *International Journal of Human-Computer Studies* 61, 751–789 (2004)
10. Su, N.M., Mark, G.: Communication chains and multitasking. In: Proceedings of the ACM Conference on Human Factors in Computing Systems - CHI 2008, pp. 83–92. ACM, New York (2008)
11. Tu, J., Tao, H., Huang, T.: Face as mouse through visual face tracking. *Computer Vision and Image Understanding* 108, 35–40 (2007)
12. Yanco, H.A.: Wheellesley: A robotic wheelchair system: Indoor navigation and user interface. In: Mittal, V.O., Yanco, H.A., Aronis, J., Simpson, R. (eds.) *Assistive Technology and Artificial Intelligence*, pp. 256–268. Springer, Berlin (1998)

Using Triaxial Acceleration Sensors as Input Devices for Disabled Persons

Matthias Söllner, Philipp Hofmann, and Josef Pösl

University of Applied Sciences Amberg-Weiden,
Faculty of Electrical Engineering and Information Technology,
Kaiser-Wilhelm-Ring 23, D-92224 Amberg, Germany
philipp.h@vr-web.de, {m.soellner, j.poesl}@haw-aw.de
<http://www.haw-aw.de>

Abstract. We propose the use of triaxial acceleration sensors as input devices for disabled persons. The use of accelerometers for interacting with computer systems is explained. Different possible applications are shown. Experiments with a commercially available sensor module are presented. Problems are discussed and future improvements are suggested.

Keywords: triaxial acceleration sensor, input device, disabled persons.

1 Introduction

In our modern world we use computers for all sorts of communication. But there is a small group of people in our society who can't communicate with others because of very strong disabilities. These people are often very clear in mind, are able to see (and/or hear), but can't use writing (or speech) for communication with their surrounding world. They even can't use computers for lack of suitable input devices. Some of them have the possibility to move at least a few muscles, even if they have not full control over them. We aim on a solution to detect single movements of different parts of the human body in order to give these people a possibility to interact with computers and communicate with their personal carer. For this purpose we propose the use of triaxial acceleration sensors, which are recently available in modern electronics.

2 Triaxial Acceleration Sensors

Nowadays many manufacturers of integrated circuits offer sensors for measuring accelerations based on MEMS (MicroElectroMechanical Systems) technology. By using micromechanical structures very small spring-mass systems are realized [1]. They are manufactured in silicon wafers by micromachining technologies like material deposition, patterning and etching. The moving mass and the springs consist of silicon stripes only a few microns thick. The position of the mass,

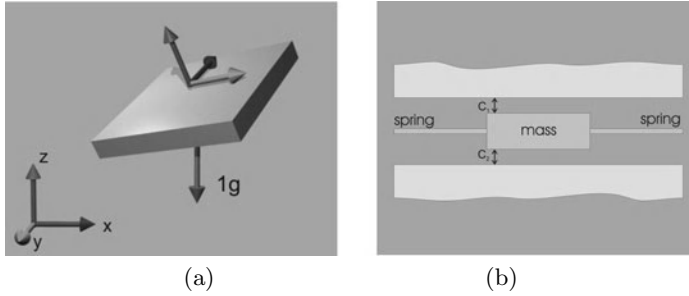


Fig. 1. Triaxial sensor (a), schematic of 1D-sensor (b)

which relates to the actual acceleration, is measured by evaluating the differential capacitance between the mass and its surrounding. The generated signals are amplified on the same chip by an integrated circuit. The measurement values are available on external pins of the sensor either as analog or digital signals. Recent innovations in micro manufacturing allow the production of triaxial sensors which can measure the accelerations in all three perpendicular directions of the space (x-y-z, Fig. 1) simultaneously. For each axis one output signal is generated. Due to the acceleration of gravity the sum of the three outputs is always $9.81 \frac{m}{s^2} \hat{=} 1g$ if the sensor is at rest.

3 Using Triaxial Acceleration Sensors as Input Devices

For using this type of sensor as input device, it is necessary to place the sensor at a part of the body which still can be controlled by the disabled. Even if the motor function is strongly handicapped and only small movements (jerks) are possible, the sensor can be placed at these parts of the body in order to detect motion. To successfully evaluate the motion signals it is sufficient if the user is able to lightly shake or nod his head, lightly move one leg or arm, or to move one finger or toe. Even the possibility to move the abdomen by deep breathing might be sufficient. The direction, the strength, and the duration of the movement is not critical, because it can be evaluated and configured by software. The sensor delivers three measurement values for its acceleration in the three axes. Interesting for the purpose of motion detection is the variation of this acceleration \mathbf{a} which is physically defined as the jerk \mathbf{j} :

$$\mathbf{j}(t) = \frac{d}{dt} \mathbf{a}(t)$$

The motion detection is based on the calculation of the length of the jerk vector:

$$|\mathbf{j}(t)| = \sqrt{\left(\frac{d}{dt}a_x(t)\right)^2 + \left(\frac{d}{dt}a_y(t)\right)^2 + \left(\frac{d}{dt}a_z(t)\right)^2}$$

This calculation is carried out by a pc based software after reading the actual measurement values from the sensor. If the jerk exceeds a predefined threshold

a motion is detected and a signal is sent to the application software (Fig. 3a). The resulting binary signal can then be used as an input to various software applications. For example questions can be presented on a screen which can be answered with 'yes' or 'no'. Another possibility is to present single letters on a moving line where the user selects a desired letter by jiggling his arm thus writing complete texts. For disabled persons who are able to lightly move more muscles (e.g. an arm and two legs) more than one sensor can be used for generating different input signals (e.g. 'yes' 'no', 'up' 'down' etc.). Using a triaxial sensor gives the possibility to mount it at every part of the body independent of the mounting direction. The motion is recognised in every direction of the three-dimensional space.

4 Experiments and Results

For our experiments we used a JoyWarrior 24F8 module [2], manufactured by Code Mercenaries (Fig. 2). This ready-to-use module contains a SMB380 tri-axial acceleration sensor from Bosch Sensortec [3] and an interface to the USB port of a personal computer. One standard application of the module is the measurement of seismic activities to take part in the Quake Catcher Network from Stanford University. As in this application very small movements of the earth are measured, the sensor fullfills the requirements concerning sensitivity for detecting body motions by far. The module has been integrated in a small housing, and attached to a hook-and-pile fastener in order to mount the sensor on a leg or an arm for example. This keeps the mounting of the sensor easy. Easy mounting is necessary, as often the personal carer of the disabled does not have the technical background for special sensor mounting or configuration. To test the sensor as input device, we built a software (EyEditor, Fig. 3b) which offers single letters for selection. The letters move across the screen and one of them is marked with a special color. If the length of the jerk vector exceeds the threshold because of sensor movement, a signal is generated. The selected letter is copied to the end of the previously written text, in order to build words and sentences. For our experiments we have been supported by a disabled in persistant vegetativ state who has tested our system. He has only gross motor skills on his right arm, but nothing more. Therefore we mounted the sensor on his right hand. He understood the handling of the system very quickly. After



Fig. 2. JoyWarrior 24F8 in housing (a), wrist mounted (b)

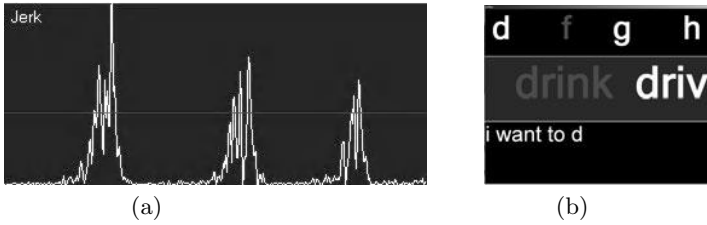


Fig. 3. Evaluation of the jerk, jiggling the hand (a), EyEditor test software (b)

a few minutes training he could write a complete sentence easily without help of any other person. This was the first self written sentence about three years after his accident. With one month training he was able to write letters of about one page length. The handling and sensor mounting is easy and can be done by the personal carer. This showed us, that using triaxial acceleration sensors has a big potential assisting disabled persons in using computers. Therefore they could give these persons the possibility to write on their own and even to communicate with other people. By using more sensors the interactivity can be increased.

5 Conclusion

We proposed using triaxial acceleration sensors as input devices for disabled persons. We could successfully show the use of a commercially available acceleration sensor module for entering texts in a computer system and giving disabled persons a chance to communicate with their environment. The sensors can be used to detect motion of arms, legs, head and other parts of the human body, thus interacting with a computer system by lightly moving some muscles. The main advantage of using triaxial sensors lies in the direction independency of movement and mounting position. In future work we will try to extend our demonstration system to use more than one sensor. We also aim on controlling mouse movements and clicks by using three or four sensors together.

References

1. Qu, W., Wenzela, C., Gerlach, G.: Fabrication of a 3D differential-capacitive acceleration sensor by UV-LIGA doi:10.1016/S0924-4247(98)00377-X
2. Code Mercenaries GmbH: JW24F8 3D acceleration sensor, http://www.codemercs.com/uploads/tx_sbdownloader/JW24F8_Datasheet.pdf (accessed January 15, 2009)
3. Bosch Sensortec GmbH: SMB380 Triaxial acceleration sensor, http://www.bosch-sensortec.com/content/language1/downloads/SMB380_Flyer_Rev1.3.pdf (accessed January 15, 2009)
4. Fukuda, Y., Tanaka, M., Yoshimochi, A., Ishimatsu, T.: Communication device to use an acceleration sensor for the seriously disabled. In: Artificial Life and Robotics, vol. 8, Springer, Japan (2004)

Canceling Interference Caused by Tremors in Joystick Controller: Study Case in a Power Wheelchair

Ludmila Correa de Alkmin Silva, Fernanda Cristina Corrêa,
Geraldo Gonçalves Delgado Neto, Vivianne V. Delgado,
and Franco Giuseppe Dedini

Department of Mechanical Engineering, Mechanical Engineering Faculty,
State University of Campinas (UNICAMP), São Paulo, Brazil
{ludmila,fernanda,geraneto,vivianne,dedini}@fem.unicamp.br

Abstract. People with disabilities such as Parkinson's, Holmes or other disease have difficulty operating conventional joysticks because of the tremor. This paper will present a study of the joystick control for the minimization of this tremor. For this minimization, different types filter and control were study and compared. These controls were tested in a wheelchair laboratory to see the behavior of the wheelchair with different inputs. Simulation results were presented to show the tremor cancel as well as the performance of the control development. With this results is possible to develop new products for people with special needs and to develop better controls for people with tremor hand.

Keywords: wheelchair, dynamics, vehicle model, tremor, control.

1 Introduction

Wheelchairs are an integral part of the lives of people with different kinds of motor impairments and the joystick is the primary control interface between a person with disability and an electric powered wheelchair [1] [2]. Unfortunately, some users of the power wheelchair have some disease that makes difficulty to operate a conventional joystick because of the vibration in the movements, mainly in the hand. These involuntary vibrations are known as tremor.

The normal tremor, commonly known as physiological tremor, shows very small amplitudes and high frequencies (greater than 8Hz) and does not represent a serious problem for most of the daily activities in human life. Although, the pathological tremor is much slower, occurring in frequency of 4 - 7 Hz (half of normal), and oscillation amplitude is also significantly higher causing very uncomfortable distortion to purposeful movements [3].

The purpose of this study is to develop a joystick control interface for power wheelchair, providing safe and an effective control to people with several disabilities, improving precision in human-machine control. Thus to cancel or minimize the tremor was implemented different types of filters in *Matlab*® and was tested

in a virtual laboratory for wheelchair [2]. Combining the virtual lab with the control of joystick was possible to visualize the wheelchair behavior. The result of this work is extremely important in new future studies and with the knowledge of the tremor is possible to develop a new and better joystick or other products to help people with special needs providing a safe and effective control for electric power wheelchair.

2 State of the Art

In recent works, there have been studies in the use of signal filtering for tremor attenuation, primarily dealing with pathological tremor [4] [5]. However, there are innumerable types of filters, they can be linear or non-linear depending of the system. Thus knowledge about the filter is necessary to properly design them for a specific application.

Most of the studies in this tremor area involves a finite impulse response linear equalizers trained on tremor recordings [4], or linear low-pass filtering, or bandstop filtering approaches [5], which aim to attenuate the full frequency band of tremor, while passing frequencies below 1 or 2 Hz, which are assumed to be voluntary. Also others linear filters are successful in attenuating tremor in many applications such as the adaptive filter and fuzzy filter [6], [7], [1].

Recently, Riviere and Thakor [7] have implemented a notch adaptive filter to improve the skills of people with tremor disability when using computer pen inputs. Riley and Rosen [5] have shown that using low-pass filters in joystick-based systems, can sometimes improve tracking performance of people affected by intention tremor.

Riley and Rosen [5] also have shown that in some cases, depending on the cutoff frequency and the specific subject, low-pass filters can degrade or have no effect on performance. Their final conclusion claims the need for subject-customized filter design. Diciano's work [13] have evaluated the performance of the filter (WFLC - Weighted frequency Fourier Linear Combiner) applied in wheelchair joysticks, but unsatisfactory results have presented.

Another work in this area is the mouse adapter. A assistive adapter is inserted between the mouse and the computer providing a digital motion-smoothing filtering, rejection of inadvertent mouse button clicks. This adapdter is based on an imaging stabilization used in some camera and is designed to work with any computer and operationg system.[10]

3 Tremor and Joystick Signal

For the implementation of the tremor was used the *Matlab*® program, where the tremor signal was considered as a noise signal, sinusoid signal. The test was made using the frequency variation between 2 Hz - 12 Hz, to cover the major tremors since the tremor due to the Holmes disease varies from 2 Hz - 5 Hz, Parkinson's range from 3 Hz - 6 Hz, essential tremor is 4 Hz - 10 Hz and physiological tremor varies from 8 Hz - 12 Hz [8].

Normally in power wheelchair are used two types of joystick: digital and analog joystick. The signals generated from these joysticks are added with the tremor noise and the resultant signal passes through the filter or another system to eliminate the noise, as can be seen from figure 1.



Fig. 1. Flowchart of signal of the joystick system

Figure 2 shows the signals adopted for the controls tests. Two types of signal were used one to represent the digital joystick and another the analog joystick. In gray line with an asterisk is the desired signal, in gray line is the tremor noise and the black is the real signal, which is the desired signal plus the noise. Figure 2-A shows the signal for digital joystick and in B shows for analog joystick.

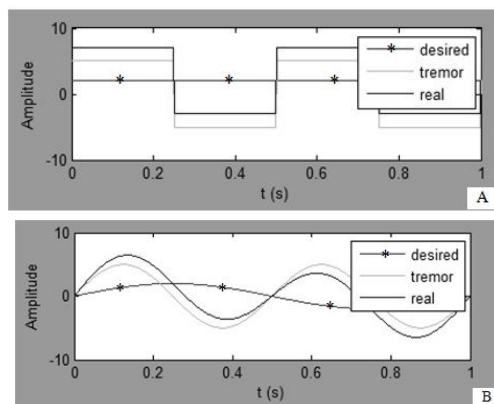


Fig. 2. Error signal (Input signal): A - Digital input and B - Analog Input

4 Tremor Cancelling Models

The first study in cancelling tremor was the mean of the input signal. The mean filter is the simplest type of low-pass filter. Mean filtering is a simple, intuitive and easy to implement method and it is used to also to reduce noise in images.

A second study was using other types of low-pass filter. Low-pass filters are filters that pass low-frequency signals but attenuate signals with higher frequencies than the cutoff frequency. The *Matlab*® presents different types of low-pass as the Elliptic, Butterworth and Chebyshev. The Butterworth filter has a

smooth passband response and more gradual out-of-band attenuation, whereas the Chebyshev filter has a ripple passband response and a steeper out-of-band attenuation. Because of the frequency and the linear response was used the Butterworth filter.

Figure 3 shows the output signal with the mean and the low-pass filter for the inputs presented in figure 2. From simulation was noticed that in the frequency between 2 Hz and 3 Hz the low-pass filter doesn't work, but with higher frequencies it works but suffers phase delay. The idea of applying low-pass filters is motivated by the fact that tremor is inherently characterized by frequency components higher than those of the desired movement. However, there is not an a priori reason why the best filter has to be low-pass. Although low-pass filters can smooth tremorous signals, their effectiveness relies on the assumption that all the high frequency components are undesirable [4].

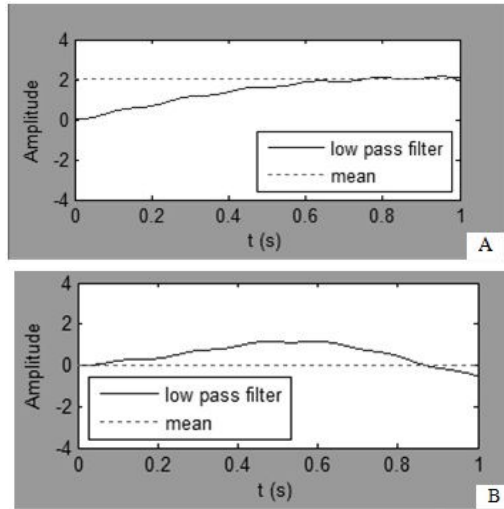


Fig. 3. Output signal using the mean and a low pass filter for a tremor with 6Hz: A - Output of digital signal and B - Output of analog signal

Another filter studied was an adaptive filter. An adaptive noise canceller is a filter that self-optimizes online through exposure to an input signal, adjusting its parameters according to a learning algorithm [9] and has an added benefit because tremor is not always constant. The advantage of the adaptive filter is to suppress only the tremor frequency and reduce distortion of intentional signals.

The principle of cancellation is to obtain an estimate of the interfering signal and subtract that from the corrupted signal. In this principle are two input to the canceller: primary and reference. The primary input is the corrupted signal, is the desired signal plus the noise, the reference input is originated from the noise [11]. Figure 4 shows an adaptive filter, it can be seen that the output

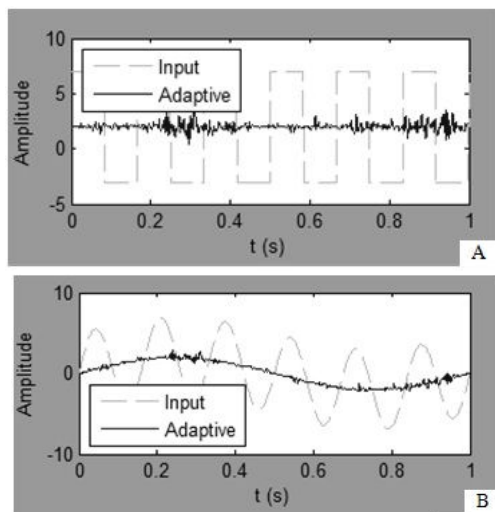


Fig. 4. Output signal using the Adaptive Filtering with 6Hz tremor: A - Digital and B - Analog

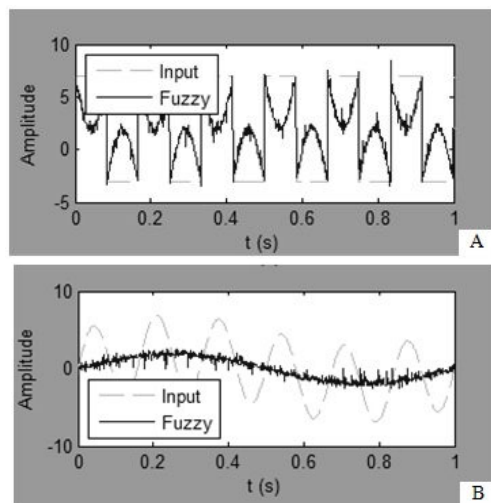


Fig. 5. Output Signal using the Fuzzy Filtering with 6Hz tremor: A - Digital and B - Analog

signal was similar of the desired input showed in figure 2. In this case was used as a reference noise a similar signal of the tremor noise for each, but if adopted another reference noise the output signal will be different.

And lastly was studied the Fuzzy Filter. Fuzzy logic rules are applied to suppress erratic hand movements and extract the intended motion from the

joystick. The procedure to obtain the fuzzy adaptive model of the adaptive fuzzy is composed of the four steps. The first is analyze the system to be modeled to understand its behavior, the second is obtain an input/output training dataset from a wellknown model, the third is define the membership functions type to be used, the fourth is define the inference method and last is optimize the fuzzy model obtained using the membership functions to improve fuzzy model performance. [12]

Figure 5 shows the simulation of fuzzy filter, in the both cases was used the same rules. The rules were generated in function of the noise tremor. In digital output can be noticed that the signal in the edge of the rectangle appears a sinusoid, this happened because was used the sinusoid noise to make the rules. This fuzzy model was obtained using the Matlab Fuzzy Logic Toolbox.

5 Simulation and Results

Attaching a joystick on a computer was possible to test the filters studied. For this test was used a Microsoft Joystick which allowed simulation of the two types of joystick, the analog and digital.

Figure 6 shows the positions of the joystick with and without low-pass filter, these positions are represented by circles with different marks. The discontinuous groove ball is the position fix, in digital joystick was developed a control that fix the desired input for some seconds. The black circle is the position after the low-pass filter and the continous groove is the real position of the joystick. As can be seen in figure 6, the tremor in the y direction, in the both simulation, the black circle stay between the position varieties and in the digital joystick the discontinuous circle stay in the desired position. These joysticks were implemented in the virtual lab of the wheelchair as showed in figure 7.

During the simulation, the wheelchair follows the commands made from the joystick, but was noticed some delay in response of the wheelchair. This may have happened because of the delay of the filter itself or because of the dynamic behavior of the own wheelchair. In the graphic of the force inputs can be seen that the oscillation tends some position as was emphasized in the figure 7 by a rectangle.

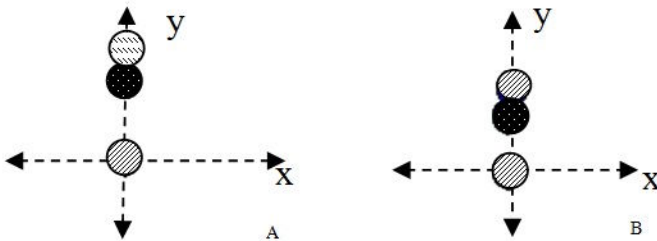


Fig. 6. Joystick test applying a tremor. A - Digital and B - Analog joystick.

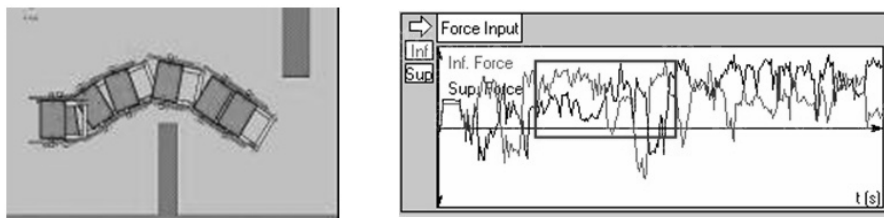


Fig. 7. Simulation Test of Analog Joystick



Fig. 8. Simulation Test of Digital Joystick

The simulation in figure 8 shows a wheelchair following the commands coming from the joystick and shows elimination of tremor in some frequency. In the graphic of the force inputs can be seen that the oscillation tends a constant in some times.

6 Conclusion

The study has promising results in cancelling tremor applied in joystick. With the help of the virtual lab was possible to visualize the behavior of the wheelchair using the controller joystick. The low-pass filter eliminated the tremor and it is also easy to be implemented in a microcontroller than other as the adaptive and fuzzy filter. Just in some case with low frequency the result of the low-pass filter was not satisfactory. And with some filter, as adaptive and fuzzy, the joystick parameters should adjusted case by case to extract intended motion from the joystick.

In the cases of low frequency should be made a deeper study to eliminate this tremor through a mechanic way instead of filters and this mechanical form can be a slack in the lever of the joystick or a new concept of joystick taking into account ergonomic factors. Thus the future work includes the ergonomic design and experimental verification. Consequently, this paper is an important contribution regarding mobility and independence of people with special needs.

Acknowledgments. The authors wish to thank CAPES, DAAD and CNPq for support this work.

References

1. Ding, D., et al.: Optimized joystick controller. In: 26th Annual International Conference of IEEE EMBS, San Francisco, US, pp. 4881–4883. IEEE Press, San Francisco (2004)
2. Silva, L.C.A., et al.: Modeling a Hands-Free Controlled Power Wheelchair. In: Miesenberger, K., Klaus, J., Zagler, W.L., Karshmer, A.I. (eds.) ICCHP 2008. LNCS, vol. 5105, pp. 1261–1268. Springer, Heidelberg (2008)
3. Gonzalez, J.G., et al.: A customized optimal filter for eliminating operator's tremor. In: Proceeding SPIE's International Symposium on Intelligent Systems and Advanced Manufacturing-Telemanipulation and Telepresence Technologies II, Philadelphia (1995)
4. Gonzalez, J.G., et al.: A new approach to suppressing abnormal tremor through signal equalization. In: Proceeding RESNA Annual Conference, pp. 707–709, Vancouver (1995)
5. Riley, P., Rosen, M.: Evaluating manual control devices for those with tremor disability. *J. Rehab. Res.* 24, 99–110 (1987)
6. Oppenheim, A.V., Schaefer, R.W.: *Discrete-time Signal Processing*, vol. 24, pp. 99–110. Prentice-Hall, Englewood Cliffs (1989)
7. Riviere, C.N., Thakor, N.V.: Assistive computer interface for pen input by persons with tremor. In: Proceeding RESNA International, Vancouver (1995)
8. Cichaczewski, E., Cunha, J.C.: Sistema computacional de auxilio ao diagnostico e aviliao de tremores. In: XV Congresso Argentino de Bioengenharia: 027EI, Argentina (2005)
9. Widrow, B., Stearns, S.D.: *Adaptive Signal Processing*, ch. 12. Prentice-Hall, Englewood Cliffs (1985)
10. Levine, J.L., Schappert, M.A.: A mouse adapter for people with hand tremor. *IBM Systems Journal* 44, 621–628 (2005)
11. Farhang-Boroujeny, B.: *Adaptive Filter Theory and Application*. John Wiley & Sons, Chichester (1998)
12. Jurez-Hermndez, M.A., et al.: A Fuzzy Adaptive Filter For System Identification. In: Proceedings of the 19th International Conference on Electronics, Communications and Computers (2005)
13. Dicianno, B.E., et al.: Joystick use for virtual power wheelchair driving in individuals with tremor: Pilot study. *Journal of Rehabilitation Research e Development* 46, 269–276 (2009)

Developing Rehabilitation Robots for the Brain Injured

Paul Gnanayutham¹ and Jennifer George²

¹ School of Computing, University of Portsmouth, Buckingham Building,
Lion Terrace, Portsmouth, PO1 3HE, United Kingdom

paul.gnanayutham@port.ac.uk

² SAE Institute, 297 Kingsland Road, E8 4DD London, United Kingdom

jennifer.george@sae.edu

Abstract. Although rehabilitation robotics have been used for helping disabled persons in various areas of disability, such as stroke rehabilitation, very little research has been done with the brain injured persons and robotics. This paper discusses the implementation of a simple model, which consists of brain body interface, a computer, an interface program and an electronic circuit to interface the computer to a robotic arm. This was an exploratory research that was carried out that allowed a brain-injured person to do simple tasks using robotic arms. This paper also looks at rehabilitation robotics both past and present. The paper goes on to explore the new avenues available to enhance this exploratory research. In this paper, we take the brain body interface communications a step further where the brain injured persons will not only communicate but will also be able to do simple tasks.

Keywords: Rehabilitation, Robot, Brain Injury, Cyberlink™, Brain Body Interface, Brain Computer Interface.

1 Introduction

The word 'Robot' come originally from a play in 1923 with the meaning a 'slave worker' (from the 1923 play Rossum's Universal Robots). The early uses of robots were mechanical devices using gears and levers. The advent of computers and the fast and furious new technology has given the robots the capability to perform sophisticated tasks others than mundane routine jobs. Robots are in action in the Military, Health sector, Manufacturing, Space exploration, Mining etc. If asked to describe a robot many people would describe something from a science fiction, our mental images are guided more by science fiction than science reality. The reality is probably a car assembly robot rather than the evil Dalek of Dr Who. Originally most robots were used as tools in a manufacturing environment, whether spraying, welding or assembling. These are normally referred to as assembly robots. It has long been realised that, while robotics will continue to have a vital role in the manufacturing industry, the areas for growth lie in applications in "*unstructured and hostile environments*". Many applications in an unstructured environment are "service" applications. Service robots have for a while been used for tasks at home such as cleaning, cutting the lawn, etc. Also hazardous environments such as searching underwater, within a nuclear reactor, the making safe of bombs and mines.

Over the past 20 years there has been much research in the area of 'advanced robotics'. The precision of a medical robot is claimed to be even better than a skilled surgeon (www.prosurgics.com). In addition to surgery, robotic technology can be used in several ways to benefit people with disabilities. Although there are many potential benefits for the disabled user, there are also very strong requirements and compromises, which must be considered in integrating the robotic arm with the wheelchair. Robots are being used in places such as Japan for caring for the physically handicapped people. These robots do the daily routine chores thus taking the burden away from the carers and also saving a lot of money. The operators control these mobile robots through the Internet and mobile phones (Yoshiyuki *et al.*, 2000). The new trend in robotics is to control robots remote using the Internet or mobile phone.

1.1 Rehabilitation Robotics from Past to Present

Tzika and team claim (2008) that the brain has the ability to regain function through rehabilitative exercises following a stroke also that the brain is malleable, even six months or more after a stroke, which is a longer period of time than previously thought. This gives hope to people who have had strokes, their families and the rehabilitative specialists who treat those (Tzika *et al.*, 2008).

The first referenced rehabilitation manipulator was the CASE manipulator built in the early 1960, which could move the user's paralysed arm (Kim and Cook, 1985). Mid 1970 saw the next stage of rehabilitation robotics in the form of a workstation based system designed by Roesler with five degree of freedom manipulator was placed in a specially adapted desktop environment, using rotating shelf units (Roesler *et al.*, 1978). Another early workstation system was that of Seamone and Schmeisser. The arm of this system was based around an electrically powered prosthetic arm, mounted on a horizontal track. Various items of equipment (e.g. telephone, book rest, computer discs) were laid out on the simple but cleverly designed workstation table and could be manipulated by the arm using pre-programmed commands (Seamone and Schmeisser, 1986). Mason and Peizer (1978), developed the first ever robot arm mounted to a wheelchair, potentially offering much greater freedom than a workstation mounted system. The four degree of freedom arm and its novel telescoping design allowed it to reach to the floor or the ceiling (Hillman, 2003). Some more examples (Hillman *et al.*, 2002a):

- Desk based system - "*Wolfson*" robot - for the desk based activities;
- Trolley mounted system - "*Wessex*" robot - requires a carer to move a trolley-based manipulator around from one room to another;
- Wheelchair mounted system - "*Weston*" robot - shares many common components with the trolley-mounted system, but mounted to a wheelchair.

All the rehabilitation robots mentioned above needed some type of motoring ability to manipulate such as a switch or joy stick and utilise the robot for carrying out various tasks. However the brain injured quadriplegic community was not catered by the developers of the rehabilitation robots above. This paper looks at the possibility of catering for this group of brain injured quadriplegic individuals to use rehabilitation robots for similar usage. The results of an exploratory study carried out using a robotic arm and bio-potentials from the forehead of a brain injured user is used as an example of what could be done to cater for this group of brain injured quadriplegic individuals.

2 Designing for the Brain Injured

There are several areas where a robotic device might assist a disabled person, here are some examples (Hillman *et al.*, 2002b):

- Eating and drinking (Fig.1);
- Personal hygiene, such as washing, shaving;
- Work and leisure, such as handling papers, books or videos (Fig.2);
- Access, such as opening doors, operating light switches or lift buttons;
- Reaching and moving, such as reaching down to pick up an item off the floor or reaching up to get an item off of a shelf.



Fig. 1. Robot for eating and drinking



Fig. 2. Work and leisure, such as handling papers, books or videos

As for the mechanical construction of attaching a robotic extension to a wheelchair there are different requirements in comparison to either a desktop or mobile system. Mounting a robotic extension into a desktop provides a "workstation" based approach, and all the items to be manipulated are placed within a known area of reach. The mobile approach is similar since the trolley-mounted robotic extension is intended to interact with objects at a number of different workstations (Hillman *et al.*, 2002b). Another choice of construction will be an agent architecture in which the agent must

operate without hindering the user's ability to take direct action when he/she will choose (Hillman *et al.*, 2002b).

3 Robots for Quadriplegic Brain Injured

Not all users with special needs can use a mouse, trackball, and keyboard or have the ability to speak to a speech recognition system. So we need a device that provides communication capabilities for those who cannot use any of the regular input devices (Gnanayutham and George, 2007 and 2009) such as:

- HeadMouse™ - using wireless optical sensor that transforms head movement into cursor movement on the screen;
- Tonguepoint™ - a system mounted on mouth piece;
- Eye Tracker™ - a system that allows the monitoring of both the conscious and unconscious gaze

All the devices above have their advantages and disadvantages. A user with cerebral palsy will not have good motor abilities to operate the 'Tonguepoint™'. A user with spinal vertebrae fusion may not be able to turn his or head and the HeadMouse™ will be of no use to this user.



Fig. 3. Cyberlink the Brain Computer Interface used in this research

At present one of the brain computer interface devices that can cater for the quadriplegic brain injured individual is the Cyberlink™ (Fig. 3) because it uses a combination of bio-potentials. 'Cyberlink™' can be used as a technology that combines eye movement, eye blink, facial muscle and brain wave bio-potentials detected at the user's forehead to generate a mouse input that can be used for communicating. Cyberlink™ uses the forehead as non-invasive site, for convenience and also because it has a rich variety of bio-potentials. The signals for communications are obtained by attaching probes on the forehead of the patients. It is 3 silver/silver chloride contact electrodes (i.e. non-invasive), which are placed on a headband that picks up EEG (brain wave), EMG (muscle movement wave) and EOG (Eye ball movement) signals when applied on the forehead. These are then fed into an amplifier box and then to the mouse port, so the computer just sees the device as a mouse, which, is used to control the cursor. The main signals used are due to muscle movement, only about

10% is due to thought processes (Brain wave). We used the Cyberlink™ to communicate with the brain injured persons to get basic yes/no answers (Gnanayutham and George, 2009a). This time we want to go a step further and make the brain injured user perform simple tasks using a robotic arm (Doherty *et al.*, 2003, Gnanayutham *et al.*, 2001).

The model for operating the robotic arm using the brain body interface consisted of following components:

1. A Cyberlink™ brain body actuated control technology system that connects to the computer via the serial port;
2. A computer with a parallel port and serial port free. An Interface program written in Visual Basic™ to operate the functions of the robotic arm;
3. An Electronic circuit to read the parallel port of the computer and operate the motors that manipulate the robotic arm;
4. A robotic arm (Super Armatron™) that is operated using a series of motors.

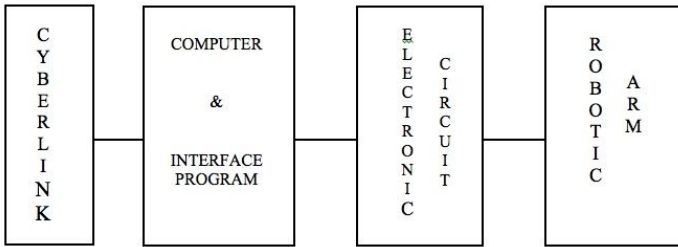


Fig. 4. Model for operating the robotic arm



Fig. 5. Interface Program

The Fig. 4, shows the model for operating the robotic arm using the brain body interface. The computer needed one serial port for the Cyberlink™ and a parallel port for the electronic circuit that interfaced the computer with the robotic arm. The Cyberlink probes were attached to the forehead of the user and the other end of the Cyberlink was connected to the serial port of the computer (Gnanayutham *et al.*, 2001).

The computer had a program written in Microsoft Visual Basic 6.0™, which had six paths for controlling the robotic arm (Fig. 5). The paths ended up in one of these functions, arm go up, arm down, arm left, arm right, open claw and close claw. When one of these six functions were triggered, the program sent a binary code to the parallel port, which drove one of the motors to carry out what was requested by the user. The Electronic circuit used in the above setup is shown below.

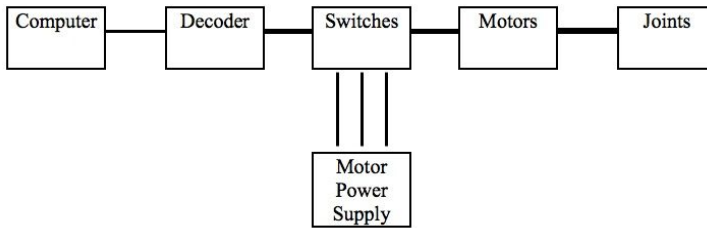


Fig. 6. Electronic Circuit

In Fig. 6, we see the block diagram of the electronic circuit that was used. The output from the parallel port was decoded and used for switching transistors. The transistors switched the motors on and off in either direction. The mechanical side of the circuit included aligning shafts and making sure there were no vibrations.

The robotic arm described in this section was developed and used as a live demonstration at the ICCIT'2001 conference in New Jersey where a quadriplegic individual picked up a cup to his mouth to show, how we can use a brain body interface and operate a robotic arm. Since then the researchers have concentrated on mainly communication, recreation and controlling the environment using brain body interfaces (Gnanayutham *et al.*, 2009b). Now the researchers feel it's time to help this group of the disabled who perhaps need robotics than other motor impaired individuals. There is one previous successful research carried out by Felzer, and Freisleben (2002) but no commercial product has come out of that piece of research. At present work is carried out at University of Essex also on navigating a wheel chair using bio-potentials (Chen *et al.*, 20007, Lai *et al.*, 2009).

4 Summary and Conclusions

This paper indicated the various choices of rehabilitation robots that are available in the market, used mainly for stroke rehabilitation and other disability where the user is able manipulate a mouse, switch or a joystick. This paper also tackled the area of giving this sophisticated and powerful robot as a tool for the brain injured who

perhaps need it more than any other category of people. Robots have been in science fiction for many years and but now we can extend these robots to the brain injured quadriplegics personnel by using their bio-potentials as the input instead of a mechanical switch, mouse or a joy stick. This exciting new research is going to change many a brain-injured person's life and set them on the path of rehabilitation.

References

1. Chun, S.L.T., Pei, J., John, Q.G., Huosheng, H., Kui, Y.: EMG-based Hands-Free Wheelchair Control with EOG Attention Shift Detection. In: Proceedings of the 2007 IEEE International Conference on Robotics and Biomimetics, Sanya, China (2007)
2. Doherty, E., Wong, Y., Fernandes, J., Gnanayutham, P., Stephenson, G.: The OWI-007 Robotic Arm, Cyberlink Interface, and Visual C++ Interface are a Great Help to the Handicapped and a Good Platform for Teaching Technology. Njudge.net Conference, New York (October 2003)
3. Felzer, T., Freisleben, B.: HaWCoS: the "hands-free" wheelchair control system. In: Proceedings of the fifth international ACM conference on Assistive technologies, Edinburgh (2002)
4. Gnanayutham, P., Bloor, C., Cockton, G.: Robotics for the brain injured: An interface for the brain injured person to operate a robotic arm. In: ICCIT 2001, New York (October 2001)
5. Gnanayutham, P., George, J.: Inclusive Design for Brain Body Interfaces. In: Schmorrow, D.D., Reeves, L.M. (eds.) HCHI 2007 and FAC 2007. LNCS (LNAI), vol. 4565, pp. 102–111. Springer, Heidelberg (2007)
6. Gnanayutham, P., George, J.: Communicating, Recreating and Controlling the Environment Using an Adaptive Brain-Body Interface. In: DSAI 2009, pp. 71–78. UTAD, Lisbon (June 2009a)
7. Gnanayutham, P., Cockton, G.: Adaptive Personalisation for Researcher-Independent Brain-Body Interface Usage. In: CHI 2009, April 2009, pp. 3003–3019. ACM Press, Boston (2009b)
8. Hillman, M.R.: Rehabilitation robotics from past to present - a historical perspective. In: ICORR, Daejeon (April 2003)
9. Hillman, M.R., Evans, N.M., Orpwood, R.D.: Improving the flexibility of an assistive Robot. In: Proceedings of the Cambridge Workshop on Universal Access and Assistive Technology 2002, Springer, Heidelberg (2002a)
10. Hillman, M.R., Hagan, K., Hagan, S., Jepson, J., Orpwood, R.D.: The Weston Wheelchair Mounted Assistive Robot – The Design Story. *Robotica* 20, 125–132 (2002b)
11. Lai, W., Huosheng, H., Kui, Y.: Use of Forehead Bio-signals for Controlling an Intelligent Wheelchair. In: Proceedings of the 2008 IEEE, International Conference on Robotics and Biomimetics, Bangkok (2009)
12. Mason, C.P., Peizer, E.: Medical Manipulator for Quadriplegic. In: IRIA 1978, New Delhi (1978)
13. Roesler, H., Kupperts, H.J., Schmalenbach, E.: The Medical Manipulator and its Adapted Environment: A System for the Rehabilitation of Severely Handicapped. In: Proc. Int'l Conf. on Telemanipulators for the Physically Handicapped (1978)

14. Seamone, W., Schmeisser, G.: Evaluation of the JHU/APL Robot Arm Workstation, Interactive Robotic Aids. World Rehabilitation Fund Monograph#37, York (1986)
15. Tzika, A., Khanicheh, A., Rosen, B., Astrakas, L., Moskowitz, M.: Stroke Rehabilitation Improved By Robotic Technology, Medical News Today (December 2008)
16. Kim, Y., Cook, A.M.: Manipulation and Mobility Aids. Electronic Devices for Rehabilitation. Chapman and Hall, London (1985)
17. Yoshiyuki, T., Takashi, K., Tateki, U., Masao, M., Hiroyuki, K.: Development of the Mobile Robot System to aid the daily life for Physically Handicapped. In: Proceedings of the 7th International Conference on Computers Helping People with Special Needs, Boston (2000)

A Rehabilitation Method with Visual and Haptic Guidance for Children with Upper Extremity Disability

Kup-Sze Choi¹, Chum-Ming Chow¹, and King-Hung Lo²

¹ School of Nursing, The Hong Kong Polytechnic University
Hung Hom, Kowloon, Hong Kong

² Department of Occupational Therapy, Hong Kong Red Cross Princess Alexandra School
Kwun Tong, Kowloon, Hong Kong
kschoi@ieee.org

Abstract. Activities of daily living present difficulties to children with upper extremity disabilities. Among the activities, handwriting is considered essential to the children, and occupational therapy is important for them to adapt to the required motor skills. In this paper, a virtual-reality based system is developed for training fine motor skills in the context of writing Chinese characters. A haptic device equipped with a pen-like stylus is employed as user interface. Haptic cues in the form of feedback forces are provided through the device to drag the child's hand towards the correct path and direction. Real-time quantitative measurements can be recorded to allow for data analysis and performance tracking.

Keywords: hand rehabilitation, motor skills, virtual reality, haptic feedback, Chinese handwriting.

1 Introduction

To children with upper extremity disabilities, hand rehabilitation plays a key role to improve their motor skills and to help them coping with activities of daily living (ADL). This is important to enhance their self-care ability and thereby creating a stronger sense of independence. Among various ADL, handwriting is considered essential to children since they may spend up to 60% of their school time in handwriting [1]. Handwriting competency is also important for children to build self-esteem and achieve success in school [2, 3]. Legible handwriting, however, is luxurious to children with upper extremity disabilities, e.g. children with motor impairment due to cerebral palsy.

To facilitate the training of fine motor skills required for handwriting, this paper presents a virtual-reality (VR) based rehabilitation method in the context of writing Chinese characters. Instead of conventional computer mouse or keyboard, natural user interface is employed for children to practise handwriting by holding a pen-like haptic device. Haptic cues in the form of feedback forces are provided through the haptic device to drag the child's hand towards the correct path and direction. The children's performance is recordable automatically in real time. Since a generic personal

computer and a low-cost desktop haptic device are only used, it is anticipated that the proposed method can become an affordable rehabilitation approach. The interactive system enables self-learning with minimal guidance from teachers or occupational therapists, making it promising to reduce caregivers' workload. Besides, the computer-game-like settings can motivate the children to participate in rehabilitative training.

2 Related Work

VR technology has been capitalized to develop new training methodologies and assessment methods for rehabilitation purposes [4, 5]. For example, many VR applications have been developed for hand rehabilitation after stroke [6, 7] and brain lesions [8]. Applications developed for handwriting training [9-11] and assessing motor abilities of upper limbs [12] have also been realized. In the system developed by Mullins et al., user's hand is controlled through the motions to form English characters but then the user plays a rather passive role in the process [9]. Pernalet et al. use their system to improve English handwriting indirectly, where the children do not perform handwriting but virtual labyrinth tasks with the system [10]. The target users of Teo's Chinese handwriting training system are people without upper extremity [11]. While stereographic display is provided, it indeed complicates the training task and presents extra difficulty for disabled children with cerebral palsy.

In these applications, haptic devices are employed for the benefit of rehabilitation since the sense of touch is an important perception channel. Research has shown that hand-eye coordination is improved by the presence of force feedback in rehabilitation [13]. Force feedback is also found to be helpful to disabled people whose motor skills are lost partially but their sense of touch remained [14]. The haptic devices are able to measure the position and orientation of user's movements continuously during rehabilitative training while feedback forces are provided at the same time [7, 12, 15]. Kinematic and dynamic motor abilities of upper limbs can thus be tracked quantitatively for objective performance evaluation [12, 16]. Furthermore, a versatile system capable of conducting several manual dexterity assessment methods adopted in occupational therapy has been implemented with haptic devices [14].

3 The Virtual Handwriting Trainer

The target users of the proposed VR rehabilitation system are children with upper extremity disability. It is used for training fine motor skills in the context of writing Chinese characters. The major hardware involved is a generic personal computer system and a haptic device equipped with a pen-like stylus. A virtual environment is created in which children are guided by visual and haptic cues to practise handwriting. The children write with a virtual pen by manoeuvring the stylus of a haptic device and watch the formation of the corresponding characters on the computer screen. The position of the virtual pen tip is recorded continuously and automatically by the system. The system simulates the scenario of real handwriting on paper so that it is natural and intuitive for virtual writing, where the tip of the stylus is moved horizontally to make contact with a virtual flat planar surface.

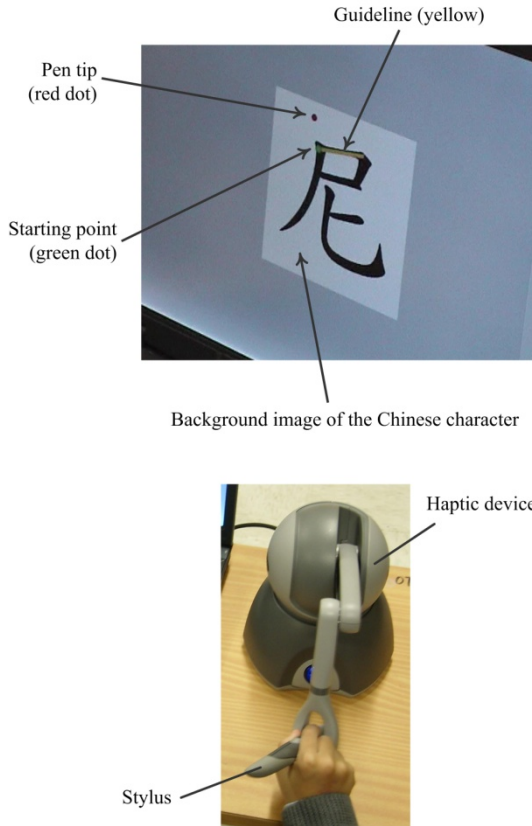


Fig. 1. Handwriting training guided by visual and haptic cues

The visual and haptic cues provided by the virtual environment are generated in real time in response to the actions of the children. As shown on the top of Fig. 1, visual cues in the form of green dot and yellow line are overlaid on each stroke of the character to indicate respectively the starting point of a stroke and the correct path to draw the stroke. A red dot is used to indicate the position of the virtual pen tip. At the bottom of Fig. 1, the child maneuvers the haptic device's stylus in a way similar to holding a real pen. The tip of the stylus corresponds to the tip of the virtual pen, which is displayed on the screen as a red dot. Feedback forces are transmitted to and perceived by the child's hand through the stylus. These forces include the reaction force exerted on the hand by the virtual planar writing surface, so that the pen tip does not go below the surface. Artificial forces are also computed on the fly to guide the user in the course of virtual handwriting. User performance can be assessed quantitatively and objectively by referring to the timing data of character writing, as well as the location and locus of the virtual pen tip.

The proposed method has been implemented with a 17-inch Hewlett-Packard notebook computer with Intel Core 2 Duo T7500 2.2GHz CPU, 2G DDR2 RAM and

Nvidia Quadro FX 1600 512MB display card. The haptic device employed is the Phantom[®] Omni[®] manufactured by the SensAble Technologies Incorporated. The device has 6 degrees-of-freedom input (position and orientation of the stylus endpoint) and 3 degrees-of-freedom output (feedback force). In this study, only the position data of the stylus are used as inputs to the software.

3.1 Visual Cues

Chinese character is composed of strokes. To write a character, the strokes are drawn in a specific sequence to form the character. In the proposed method, training begins by showing the image of the complete character at the background. For each stroke, yellow guideline is displayed and superimposed on top of the background image. The yellow guidelines, displayed sequentially according to the stroke order, serve as visual cues to hint and guide the user to draw the strokes one by one in the correct sequence. The situation is shown in Fig. 2. Note that the figure illustrates a simple case where the guidelines consist of one single line segment only. In general, the guideline for drawing a stroke may be defined by a *concatenation of line segments*, especially for curly stroke. The example given in Fig. 3 demonstrates that a guideline of two line segments is required for drawing the first stroke of the Chinese character "good".

To begin drawing a stroke, the virtual pen tip (red dot) is moved to make contact with the starting point (green dot). After that, the child is supposed to move the red dot along the yellow guideline until reaching the other end to emulate real handwriting.

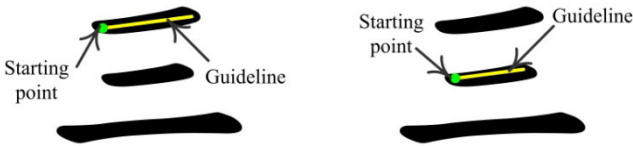


Fig. 2. Visual cues to guide the writing of first (left) and second stroke (right) of the Chinese character "three"

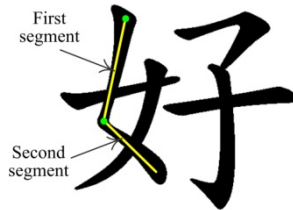


Fig. 3. The two-segment guideline for the first stroke of the Chinese character "good"

3.2 Defining the Guidelines

The guideline for each stroke of a character is obtained by manually identifying the coordinates of the starting point and the end point of the constituting line segments for each guideline. The segments are defined sequentially according to the stroke order and the way that the strokes are drawn, which are then encoded and stored in a file. To facilitate this manual process, a software widget is developed to enable the definition of appropriate segments of the guideline for drawing a stroke, by using a computer mouse to click at the end points the required guidelines. Following the stroke order, the coordinates associated with the guidelines for individual strokes are stored in sequence and saved as an XML file. Refer to the character shown in Fig. 3. It is composed of six strokes and the corresponding XML file defining the stroke orders and the coordinates of the associated guidelines are shown below. The two segments of the guideline for the first stroke are defined in lines 3-5 of the XML file. Similarly, the guidelines for drawing the second, fourth and fifth stroke also contains two segments, while the other guidelines contain one segment only.

```

1.    <Character strokes="6">
2.    <Stroke id="0" points="3">
3.        <Point id="0" x="3.200000" y="8.100000" />
4.        <Point id="1" x="2.500000" y="4.400000" />
5.        <Point id="2" x="4.500000" y="2.500000" />
6.    </Stroke>
7.    <Stroke id="1" points="3">
8.        <Point id="0" x="4.300000" y="6.400000" />
9.        <Point id="1" x="3.800000" y="4.000000" />
10.       <Point id="2" x="2.200000" y="2.400000" />
11.    </Stroke>
12.    <Stroke id="2" points="2">
13.        <Point id="0" x="1.500000" y="5.100000" />
14.        <Point id="1" x="5.000000" y="5.900000" />
15.    </Stroke>
16.    <Stroke id="3" points="3">
17.        <Point id="0" x="5.300000" y="7.500000" />
18.        <Point id="1" x="7.700000" y="8.000000" />
19.        <Point id="2" x="6.700000" y="6.500000" />
20.    </Stroke>
21.    <Stroke id="4" points="3">
22.        <Point id="0" x="6.500000" y="6.300000" />
23.        <Point id="1" x="6.600000" y="1.800000" />
24.        <Point id="2" x="5.800000" y="2.300000" />
25.    </Stroke>
26.    <Stroke id="5" points="2">
27.        <Point id="0" x="4.900000" y="4.900000" />
28.        <Point id="1" x="8.800000" y="5.400000" />
29.    </Stroke>
30. </Character>

```

3.3 Haptic Cues

In the course of guided writing, haptic cues are also provided simultaneously. Forces are generated to assist the user by forcing the pen tip to move along the guidelines. If the pen tip deviates from the guideline, leaving the writing surface, or moving away from the guideline laterally, forces are produced to pull it back to the guideline. These forces are generated on the fly to drive the haptic device and eventually transmitted to the user. In the VR system, two kinds of haptic cues are provided. First, *dragging forces* are generated, after the user has reached the starting point, to move the virtual pen tip from the starting point of a guideline towards its end. If the user moves the pen tip away from a guideline, *guiding forces* are computed to pull it back to the guideline. The situations are depicted schematically in Fig. 4.

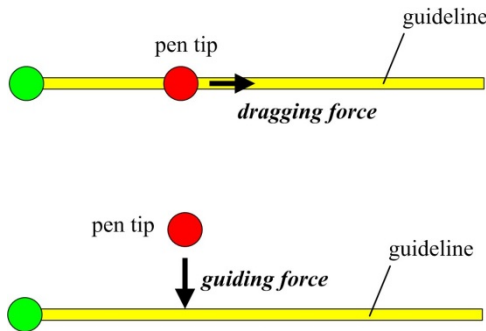


Fig. 4. Dragging force (top) and guiding force (bottom)

The difficulty of rehabilitative training can therefore be graded by removing the amount of haptic cues provided at different stages of the training. For example, both dragging and guiding forces can be provided at the beginning and then guiding forces are only made available, and finally, all haptic cues are removed to allow user to move freely on the virtual writing surface. The developed VR system can provide quantitative metrics to evaluate user's performance during the course of handwriting. The time elapsed to complete writing a character can be measured for manual dexterity analysis. Shorter completion time suggests better hand-eye coordination. The location and trajectory of the pen tip are also recorded to give a visual overview regarding how well a user can in keeping the pen tip within the guidelines when writing a character.

4 Conclusion and Future Work

The paper proposes a rehabilitative training approach for children with upper extremity disability to practise fine motor skills and hand-eye coordination through force-guided handwriting. Visual and haptic cues are generated in real time to guide the user to write in correct ways. The amount of haptic cues can be adjusted according to the degree of disability and the learning progress. Children's performance can be measured with quantitative measures and conveniently recorded for progress monitoring. The method

has the potential to serve as an assessment tool in conjunction with conventional approaches to evaluate hand dexterity in a more comprehensive manner.

A pilot study is being conducted to evaluate the feasibility of the proposed rehabilitative training method. The purposes are (i) to conduct a preliminary study to determine whether the proposed method can help improving hand dexterity of disabled children, and (ii) to identify issues and rooms for further enhancement in order to make it practical for use as a routine training rehabilitation exercise. In the study, subjects are arranged to attend several training sessions and instructed to use the proposed VR system to practise Chinese handwriting with the haptic device. Visual guidance is provided throughout the study while force guidance is graded by gradually removing the haptic cues to increase the difficulty. Their performance is assessed by making reference to the legibility of handwriting on paper with a pencil, and the amount of time the children spend on writing the test characters without force guidance. Initial evaluation shows that the subjects are interested in the VR rehabilitation method. The method is able to help the children to understand the correct stroke order in writing a Chinese character and the legibility appears to improve in some cases. Further investigation will be required to improve the design of the VR system and the training scheme.

Acknowledgments. This work is supported in part by the Hong Kong Polytechnic University (Block Grant, Project account code 1-ZV5E and 1-ZV2U). The author K.S. Choi is grateful to the support of the Principal S.L. Wu, the OT staff, and T.Y. Chan.

References

1. McHale, K., Cermak, S.: Fine Motor Activities in Elementary School: Preliminary Findings and Provisional Implications for Children with Fine Motor Problems. *Am. J. Occup. Ther.* 46(10), 898–930 (1992)
2. Feder, K.P., Majnemer, A.: Handwriting Development, Competency and Intervention. *Developmental medicine and child neurology* 49(4), 312–317 (2007)
3. Steward, S.R.: Development of Written Language Proficiency: Methods for Teaching Text Structure. In: Simon, C.S. (ed.) *Communication Skills and Classroom Success*, pp. 419–432. Thinking Publications, Eau Claire (1992)
4. Burdea, G.C.: Virtual Rehabilitation - Benefits and Challenges. *Methods of Information in Medicine* 42(5), 519–523 (2003)
5. Holden, M.K.: Virtual Environments for Motor Rehabilitation: Review. *Cyberpsychology and behavior* 8(3), 187–211 (2005)
6. Merians, A., Jack, D., Boian, R.: Virtual Reality-Augmented Rehabilitation for Patients Following Stroke. *Phys. Ther.* 82, 898–915 (2002)
7. Broeren, J., Rydmark, M., Katharina, S.: Virtual Reality and Haptics as a Training Device for Movement Rehabilitation after Stroke: A Single-Case Study. *Arch. Phys. Med. Rehabil.* 85, 1247–1250 (2004)
8. Piron, L., Cenni, F., Tonin, P., Dam, M.: Virtual Reality as an Assessment Tool for Arm Motor Deficits after Brain Lesions. *Stud. Health Technol. Inform.* 81, 386–392 (2001)
9. Mullins, J., Mawson, C., Nahavandi, S.: Haptic Handwriting Aid for Training and Rehabilitation. In: *IEEE International Conference on Systems, Man and Cybernetics*, pp. 2690–2694 (2005)

10. Pernalet, N., Edwards, S., Gottipati, R., Tipple, J., Kolipakam, V., Dubey, R.V.: Eye-Hand Coordination Assessment/Therapy Using a Robotic Haptic Device. In: 9th International Conference on Rehabilitation Robotics, pp. 25–28 (2005)
11. Teo, C.L., Burdet, E., Lim, H.P.: A Robotic Teacher of Chinese Handwriting. In: 10th Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems, pp. 335–341 (2002)
12. Bardorfer, A., Muni, M., Zupan, A., Primožic, A.: Upper Limb Motion Analysis Using Haptic Interface. *IEEE/ASME Transactions on Mechatronics* 6(3), 253–260 (2001)
13. Arsenault, R., Ware, C.: Eye-Hand Co-Ordination with Force Feedback. In: SIGCHI conference on Human factors in computing systems, pp. 408–414. ACM, The Hague (2000)
14. Pernalet, N., Wentao, Y., Dubey, R., Moreno, W.: Development of a Robotic Haptic Interface to Assist the Performance of Vocational Tasks by People with Disabilities. In: IEEE International Conference on Robotics and Automation, pp. 1269–1274 (2002)
15. Broeren, J., Bjorkdahl, A., Pascher, R., Rydmark, M.: Virtual Reality and Haptics as an Assessment Device in the Postacute Phase after Stroke. *Cyberpsychol Behav.* 5, 207–211 (2002)
16. Prisco, G.M., Avizzano, C.A., Calcarà, M., Ciancio, S., Pinna, S., Bergamasco, M.: A Virtual Environment with Haptic Feedback for the Treatment of Motor Dexterity Disabilities. In: IEEE International Conference on Robotics and Automation, pp. 3721–3726 (1998)

SSVEP Based Brain-Computer Interface for Robot Control

Rupert Ortner, Christoph Guger, Robert Prueckl,
Engelbert Grünbacher, and Günter Edlinger

g.tec Guger Technologies OG, Sierningstrasse 14, 4521 Schiedlberg, Austria
{ortner, guger, prueckl, gruenbacher, edlinger}@gtec.at
<http://www.gtec.at>

Abstract. A brain computer interface (BCI) using steady state visual evoked potentials (SSVEP) is presented. EEG was derived from 3 subjects to test the suitability of SSVEPs for robot control. To calculate features and to classify the EEG data Minimum Energy and Fast Fourier Transformation (FFT) with linear discriminant analysis (LDA) were used. Finally the change rate (fluctuation of the classification result) and the majority weight of the analysis algorithms were calculated to increase the robustness and to provide a zero-class classification. The implementation was tested with a robot that was able to move forward, backward, to the left and to the right and to stop. A high accuracy was achieved for all commands. Of special interest is that the robot stopped with high reliability if the subject did not watch at the stimulation LEDs and therefore successfully zero-class recognition was implemented.

Keywords: SSVEP, BCI, robot control.

1 Introduction

A brain computer interface (BCI) is a new way of communication between humans and computers [1]. With a BCI a person ideally does not have to make use of the common output pathways of peripheral nerves and muscles. A BCI-system provides a completely new output pathway and this is perhaps the only way a person can express himself if he/she suffers on disorders like amyotrophic lateral sclerosis (ALS), brainstem stroke, brain or spinal cord injury or other diseases which impair the function of the common output pathways which are responsible for the control of muscles or impair the muscles itself [1]. In such a case one possibility is to work with the electrical brainwaves of the person. These are measured with the well-known electroencephalography (EEG), which was primarily used for clinical purposes. For a BCI the EEG is amplified and fed into a personal computer which is under certain circumstances and with appropriate algorithms able to process the data and to give the person a new kind of communication channel.

For the proposed BCI a neurological phenomenon called steady state visual evoked potential (SSVEP) is utilized. A visual evoked potential (VEP) is an electrical potential-difference, which can be derived from the scalp after a visual stimulus, for example a flash-light. VEPs after stimuli with a frequency ≤ 3.5 Hz are called “transient”

VEPs. If the stimulation frequency is > 3.5 Hz they are called “steady state” VEPs because the individual responses overlap and result a quasi-sinusoid oscillation with the same frequency as the stimulus [2]. The goal is to detect this frequency reliably with high accuracy and furthermore to detect when the frequency is not present, thus when the person does not look at the stimulus. The later one is a very challenging task in BCI systems. The paper will introduce signal processing methods that allow answering these questions.

2 Methods

Three healthy subjects participated in the BCI experiment and performed first the training and then the testing procedure. The training procedure starts with a 20 second brake to have baseline EEG activity. Then each light is flashing sequentially for 14 s with 6 s breaks in between. This loop is repeated 3 times. The test procedure layout looks identical with the only exception that the each light is flashing three times in random order.

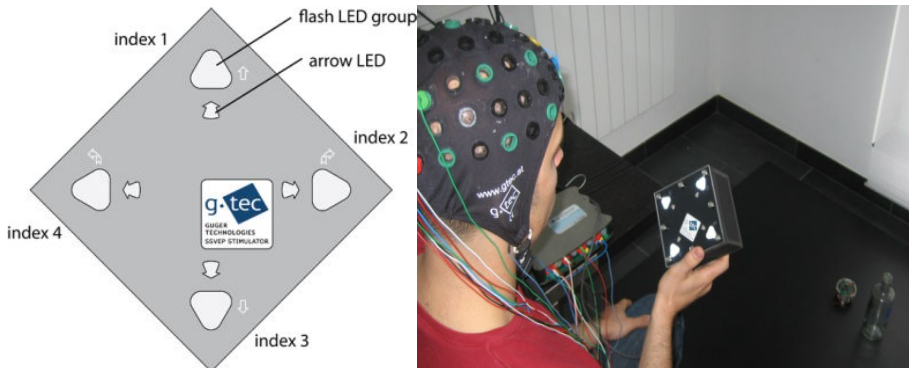


Fig. 1. The layout of the LED stimulation box and a subject wearing an electrode cap with mounted electrodes. The electrodes are connected to the biosignal amplifier g.USBamp (g.tec medical engineering GmbH, Austria). The robot is located on the right side of the picture, besides the bottle.

The SSVEP stimulation is realized with a 12x12cm box (see Figure 1) equipped with four LED-groups containing three LEDs each. Each LED has a diameter of 8 mm and according to the manufacturer a light intensity of 1500 mcd. A semitransparent foil was put over the LEDs to make them look like one compact light source. Additionally four arrow LEDs were added to indicate the index the user has to look at to train the BCI system. The LEDs are controlled by a microcontroller connected to the computer via USB. The accuracy of the produced frequencies has been validated using a digital oscilloscope. The measured maximum frequency error is < 0.025 Hz at room temperature.

The EEG-data is derived with eight gold electrodes placed mostly over visual cortex on positions POz, PO3, PO4, PO7, PO8, O1, O2 and Oz of the international 10-20

system. The reference electrode was placed at the right earlobe and a ground electrode at position FPz. Abrasive gel was applied to make sure that impedances were below 5 k Ω . The electrodes were connected to EEG-amplifier g.USBamp (g.tec medical engineering GmbH, Austria) which fed the signals over a USB connection into a personal computer. EEG data were recorded with a sampling rate of 256 Hz. The bandpass was set to 0.5 to 60 Hz and the notch filter was set to 50 Hz.

The EEG data is analyzed with several feature extraction and classification methods resulting in a classification output for each method. Each classifier has a discrete output in the form of a number (1, 2, 3 and 4) that corresponds to a certain LED. Finally in the last processing stage, the change rate/majority weight analysis step adds a 0 to this set of outputs. The device driver of the robot transforms these five numbers semantically to driving commands (0-stop, 1-forward, 2-right, 3-backward, 4-left) and sends them to robot, which moves and gives the feedback to the user.

The four LEDs are flickering with different frequencies (10, 11, 12 and 13 Hz). These frequencies have been chosen in preceding off-line tests and showed good performance for the test subjects and are also known from literature to give good accuracy [4].

All the components of the BCI system are shown in Figure 2. EEG data are recorded with a sampling rate of 256 Hz. The overall process (core system in Figure 2) works with 2-second windows (512 samples) with an overlap of 448 samples and consists of four steps: pre-processing, feature extraction, classification and change rate/majority weight analysis. These three steps are executed four times a second to have a new output every 250 ms. The paradigm controls the stimulation.

In the pre-processing step Laplacian derivations are calculated. Each Laplacian derivation is composed of one center signal X_C and an arbitrary number $n > 1$ of side signals $X_{S,i}, i = 1, \dots, n$ which are arranged symmetrically around the center signal. These signals are then combined to a new signal $Y_j = n \cdot X_C - (X_{S,1} + \dots + X_{S,n})$ where j is the index of the derivation.

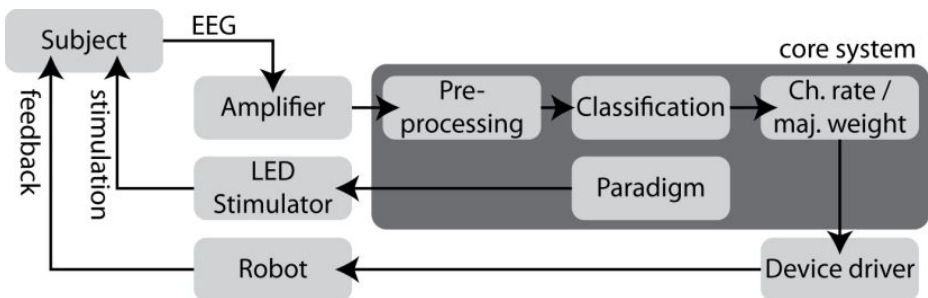


Fig. 2. BCI-system

Two different methods are used to calculate features of the EEG data. One is the minimum energy approach (ME), which was published by O. Friman et.al. in 2007 [4] and requires no training. This algorithm is fed with raw EEG-data channels since it selects the best combination of channels by itself. First of all the EEG-data gets

“cleaned” of potential SSVEP-signals. This is done by projecting artificial oscillations with stimulation frequencies and harmonics onto the orthogonal complement of the EEG-signals. After that operation the signals contain just the unwanted noise. Now a weight vector is generated, which has the property of combining the channels in a way, that the outcome has minimal energy. Now SSVEP detection is done utilizing a test statistic which calculates the ratio between the signal with an estimated SSVEP-response and the signal where no visual stimulus is present. This is done for all stimulation frequencies and all EEG-channels. The output of this classifier is the index of the frequency with the highest signal/noise ratio.

As second method a Fast Fourier Transformation (FFT) and linear discriminant analysis (LDA) using the Laplacian derivations is used. First of all the incoming data gets transformed to the frequency spectrum with a 1024-point FFT. A feature vector is extracted by taking the values of the stimulation frequencies and their 1st and 2nd harmonics. With these feature vectors a weight/bias vector must be generated for each user in a training procedure. When the training was completed successfully the LDA classifier can then be used to classify new feature vectors to one of the stimulation frequency indices. In the model used for the experiments described in this paper four ME classification units and four FFT+LDA classification units were used with different EEG channels as inputs.

The last step is a procedure called change rate/majority weight analysis. By having multiple classification units configured with slightly different input data there will be in general random classification results on noise input. This effect is used on one side to produce a zero decision when the outputs of the classifiers are changing heavily and are very different. On the other side a low change rate and a high majority weight (the number of classifications of the different algorithms which are pointing in the same direction) can be used to strengthen the robustness of the decision. Calculation is made on the last second. Default thresholds of 0.25 for change rate and 0.75 (1 – all outputs are pointing into the same direction) for majority weight were used.

The first step of the procedure is to look at the change rate. If it is above the threshold the procedure returns a final classification result of 0 which corresponds to stop the robot. Otherwise, if it is below the threshold the next step is to look at the majority weight. If this is above the threshold the majority is taken as final result, otherwise the final output is again 0. The final classification is then sent to the device controller and the robot.

3 Results

Table 1 shows the results of the testing phase. The error rate includes the whole data set which means that also the breaks were classified to test the null classification when the subject was not looking at the lights to stop the robot.

Subject 1 for example had an overall error rate of 22.7%. This error breaks down in 58.5% with no decision (robot stopped where SSVEP stimulation actually happened) and 41.5% of wrong decisions (where the chosen class was wrong, unconcerned if there had been stimulation or not)

There exists of course a delay between the target classification and the real output of the BCI. This is caused on one hand by the data collection for the 2-second analysis window of the classifiers and on the other hand by the change rate/majority analysis

section which collects four classification outputs for its calculations, thus needs 1 second. Smaller delays like the physiological delay of SSVEP itself, from perception of the stimulus until the potential is measurable, or delays between sending the “stimulation on/off” signal from the computer to the microcontroller of the stimulation box, have been unattended here. This gives in total a delay of 3 seconds. To get an idea how the error looks if this delay is disregarded the final classification result was shifted backwards for 768 samples. For online processing this would theoretically mean that at the time of analysis the time windows of the system would already be filled with appropriate EEG data. This gives a more objective view to the classification quality itself. As you can see on the right side of Table 1 subject 1 had an overall error rate of 9.5% with a fraction of 28.3% of wrong classifications. An error of 9.5% seems to be high, but it includes also the breaks between the stimulations. In total 1088 classifications were made during one run: 20 sec + (15 + 7 secs) * 3 runs * 4 LEDES * 4 decisions/sec = 1088 decisions per experiment. Out of the 1088 decisions only 28 wrong classifications were made during the whole experiment including the breaks.

Table 1. Results of SSVEP tests for 3 subjects. The error is calculated by comparing the target with the actual classification. The table shows the results without delay (target and actual classification are directly compared) and with a 3 seconds delay (the actual classification has a delay of about 3 seconds and therefore the target was shifted forward for the comparison). The overall error splits up into two subcategories of errors. No decision: no majority could be found for one class. Wrong class: the classification result was not equal to the target classification.

Subject	Without delay			With 3 seconds delay		
	Error [%] ME+LDA	No decision [%]	Wrong class [%]	Error [%] ME+LDA	No decision [%]	Wrong class [%]
S1	22,7	58,5	41,5	9,5	71,7	28,3
S2	35,7	77,9	22,1	23,5	92,7	7,3
S3	28,7	63,8	36,2	18,9	75,0	25,0
Mean	29,0	66,7	33,3	17,3	79,8	20,2

4 Discussion

A BCI system based on SSVEPs was presented which delivers a continuous command stream and has a robust zero-classification mechanism. Three subjects participated in tests and achieved an average error rate of 29%. Of these errors 66.7% on average are zero-class errors where the robot remains stopped and executed no wrong command. Thus the average percentage of wrong commands seen for the whole experiments was 9.7%. This is a great performance for controlling the movement of a robot including the zero class.

In future test runs it is necessary to evaluate other parameter configurations (source derivations, electrode positions, analysis window lengths, feature extraction procedures, thresholds for change rate/majority analysis) to optimize the error rates and the

delay. This is important for providing fast feedback to the user to give him a precise and crisp feeling of robot control.

Further tests will use a predefined route that must be followed with the robot to observe performance parameters such as duration, excursions, behaviour of the test person when looking between box and feedback of the robot, ... That would not only give an impression of the error rate, but also of the usability of the system.

It would also be very interesting to test the performance of a computer screen stimulator and compare it to the LED stimulator. In some other test runs partly other electrode positions were used which lay below position Oz. Experiments showed that this yields to a further improvement. Furthermore tests have shown that for some subjects LDA had superior performance and for other subjects ME worked better. Further experiments are needed to optimize this configuration.

Acknowledgement

This work was funded by EC projects SM4all and Brainable.

References

1. Wolpaw, J.R., Birbaumer, N., McFarland, D.J., Pfurtscheller, G., Vaughan, T.: Brain-computer interfaces for communication and control. *Clinical Neurophysiology*, 113, 767--791, (2002)
2. Paulus, W.: Elektroretinographie (ERG) und visuell evozierte Potenziale (VEP). In: Buchner, H., Noth, J. (eds.) *Evozierte Potenziale, neurovegetative Diagnostik, Okulographie: Methodik und klinische Anwendungen*, pp. 57--65, Thieme, Stuttgart - New York, (2005)
3. Lagerlund, T.D.: EEG Source Localization (Model-Dependent and Model-Independent Methods). In: Niedermeyer, E., Silva, F.L. (eds.) *Electroencephalography: Basic Principles, Clinical Applications, and Related Fields*, pp. 829--844, Lippincott Williams & Wilkins, Baltimore, (2004)
4. Friman, O., Volosyak, I., Graser, A.: Multiple channel detection of Steady-State Visual Evoked Potentials for brain-Computer interfaces. *IEEE Transactions on Biomedical Engineering*, 54, 742--750, (2007)

Text Composing Software for Disabled People, Using Blink and Motion Detection

Philipp Hofmann, Matthias Söllner, and Josef Pösl

University of Applied Sciences Amberg-Weiden,
Faculty of Electrical Engineering and Information Technology,
Kaiser-Wilhelm-Ring 23, D-92224 Amberg, Germany
philipp.h@vr-web.de, {m.soellner, j.poesl}@haw-aw.de
<http://www.haw-aw.de>

Abstract. We present a pc software to give disabled persons the possibility of writing texts without mouse and keyboard. Single letters and words are presented, scrolling over the screen. The suggested words are chosen from a self-learning database by frequency of their use. Although different kinds of binary input signals can be used, we focused on blink detection for choosing letters and words. Algorithms based on Haar Cascades are used for finding the eyes. We aim on low-cost realisation, therefore a state-of-the-art Windows based computer and a commercially available webcam is used. As an alternative input device we used an acceleration sensor for motion detection. We could successfully demonstrate quick text writing with different test persons.

Keywords: text composing, blink detection, acceleration sensor, OpenCV, disabled persons.

1 Introduction

Composing texts is one of a few essentials for disabled persons to communicate with their surrounding. Often they can't speak or write without special equipment. There are already some systems on the market composing texts by the use of personal computers. But more often they are either very inflexible in configuration or dedicated to a small subgroup of disableds with special handicaps. Sometimes special skills are still necessary, for example many existing systems try to interpret the movements of the pupil by triangulation and use a virtual keyboard for letter selection. There are many handicapped persons who can't use such systems because of their limited field of view or a malfunction of their eye axes. Therefore triangulation will not work for them. Furthermore such systems are often not affordable for the disableds or their families if they are not financed by a health insurance which is not a matter of course. In our work we focused on building a low-cost system which can be controlled only by blinking the eyes or by motion detection using a triaxal acceleration sensor. Other input devices generating binary input signals can be used, too.

2 Editor Based on Binary Input Signals

We implemented a computer software for composing texts totally controlled by one or a few binary input signals. Due to the constricted field of view of some disableds, we decided not to implement a virtual keyboard. Instead of that, our software (EyEditor, Fig. 1) suggests letters and words, scrolling over the screen comparable to a newsticker. The user is able to select letters, words or some other control actions by generating binary input signals.

As signals we used eye blinks or jerks of some parts of the body. The screen is divided horizontally in three areas. In the lower part, the composed text is shown. In the upper and middle line two scrolling tickers are displayed. They present the suggested letters (first line) and words (second line). In each line one of the items is marked with a different color and can be selected by the user. The connection of a binary input signal to either of the two scrolling lines can be configured. We implemented a simple event triggered interface in order



Fig. 1. Output window of text editor with scrolling lines

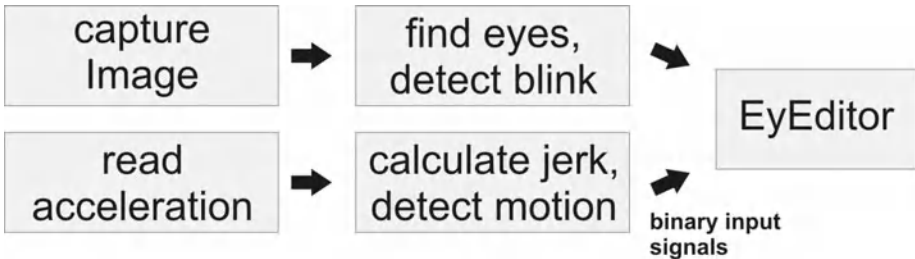


Fig. 2. Using different binary input signals

Table 1. Words in database beginning with ‘mo’

Word	Counter
motor	2
monday	4
money	2
mobile	2
mother	5

to use different types of sensors (Fig. 2). In our work we used blink detection for selecting letters [1, 2] and motion detection for selecting words [3]. Also other combinations of signals can be used. Blinking the left eye for letters and blinking the right eye for words is possible by using only one webcam. With two acceleration sensors it would be possible to select letters by tipping a toe and selecting words by nodding the head for example.

Using a database in the background the order of the suggested letters and words is based on the frequency of their occurrence. The system is self teaching as the occurrence counter of a word is increased when a user writes a word several times. The likelihood of a letter is calculated from the already written letters of a word and the frequency of occurrence in the database. For example, if a user has written the two letters ‘mo’, the database is searched for words with this beginning. The search result could look like table 1. For every word a counter in the database exists. The frequency of the third letter is summed up. In this example the letter ‘t’ has a total count of seven, therefore it is presented before the letter ‘n’ with a frequency of six. The first presented letters therefore would be ‘t’, ‘n’ and ‘b’ in this specific order. They are followed by all other letters of the alphabet with lower frequency. Letters with same rate are presented in alphabetical order. In the second line the words are sorted by their frequency (here: ‘mother monday mobile money motor’). This algorithm gives the possibility to write words and sentences very fast, because the vocabulary of the user is mapped into a database after some time of usage.

From time to time some special symbols (e.g. ‘● Δ’) are presented in the first line. If the user selects one of these symbols some special functions are shown up in the second line. The presented words in the second line are then exchanged either through special characters like punctuation marks or by control phrases like ‘Print Text, New Text, Delete Letter’. After selecting one of these options, words are suggested again. In this way the user can write special characters or trigger special functions like printing the written text.

3 Blink Detection as Input Signal

In order to generate binary input signals from eye blinks image capturing and processing is necessary. We used a cheap standard webcam for capturing the images to keep the system low-cost. To process the sequences of pictures and detect blinks, algorithms based on the open source libraries OpenCV and EmguCV are

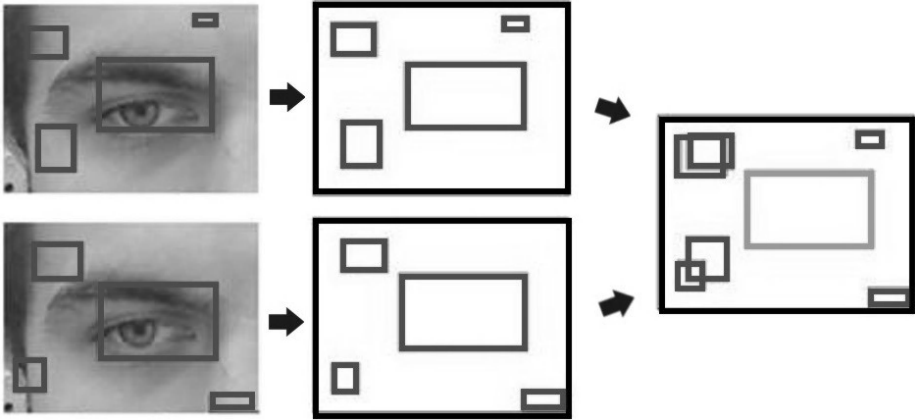


Fig. 3. Finding the eyes



Fig. 4. Finding the pupil by edge detection

used. For blink detection we first try to find a head in a picture by using Haar cascades [4, 5]. Executing this algorithm results in a few areas, which may contain a human head. To decide which area to use, the algorithm is applied to two successive frames. As they differ in image noise the algorithm generates different results for nearly the same images. The detected regions of the two frames are compared with each other. If the same area is found in both frames we use it to search the eyes.

Therefore we subdivide this region similar to the geometry of a human face and use Haar cascades on the parts where the two eyes are likely to be. The regions of the face containing forehead, mouth and chin don't have to be searched. To improve accuracy the eye finding is done on a number of sequential frames, comparing the resulting regions concerning size and local placement (Fig. 3). This results in two areas, one for the left and one for the right eye.

To find the pupil in an eye region edge detection algorithms are applied (Fig. 4) as huge differences in color between iris and sclera exist. In figure 4 found edges are marked with a circle. The edges are used to calculate the middle of the eye (the position of the pupil). For people wearing glasses, we found that template matching is more reliable, because with edge detection points laying on the glasses are often misdetections. For using template matching the software needs to be trained with images of the user's eye.

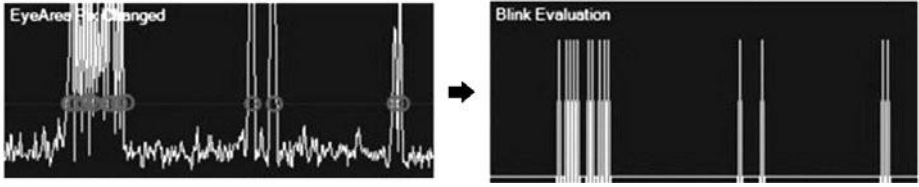


Fig. 5. Evaluation of blinks

To increase speed the search of head and eyes is only done when necessary, for example at program start or when the eye position is lost. If we already have found the two eyes, we try to track them by the Lucas-Canade algorithm as this is much faster.

Taking the actual and a past frame and comparing the intensity of the pixels surrounding the pupil, blinks are captured (Fig. 5). To distinguish an intended blinking from a moistening eyelid reflex the user has to blink twice for generating an input signal. In future we plan to distinguish also on duration of the eye closing, as this may be also a parameter to decide if it is an intended blink or not.

4 Motion Detection as Input Signal

For generating input signals by motion of the human body a triaxial acceleration sensor is used. The high sensitivity of the sensor in all three directions of the space is ideal for motion detection. The sensor has to be mounted on the specific part of the body, which the user is capable to move. This can be an arm or a leg, but also the head or even a toe. Evaluation of the jerk in all three directions generates the required trigger signal. This can be considered as an add-on but also as a replacement of the blink detection to control the Editor. For example, letters can be selected by blinking and words can be selected by moving the hand. The use of a triaxial sensor as input device is discussed in more detail in a further paper of the authors [3]. We have successfully used this technique for writing texts with our editor software.

5 Experiments and Results

We built a complete system for our tests, which is shown in figure 6. The computer system is set up on a small wagon and equipped with webcam, motion sensor and printer. The monitor is mounted on a swivel arm for easy positioning.

Different experiments for evaluating the software have been carried out. We could successfully show that entering texts with the proposed software is easy and quick. Probandes could easily write whole sentences and even complete pages without much training. Some improvements have to be done in writing numbers or complete phrases.

For evaluation of our blink detection we considered people with different hair color, hair length, face profile and some wearing glasses. Also we used different

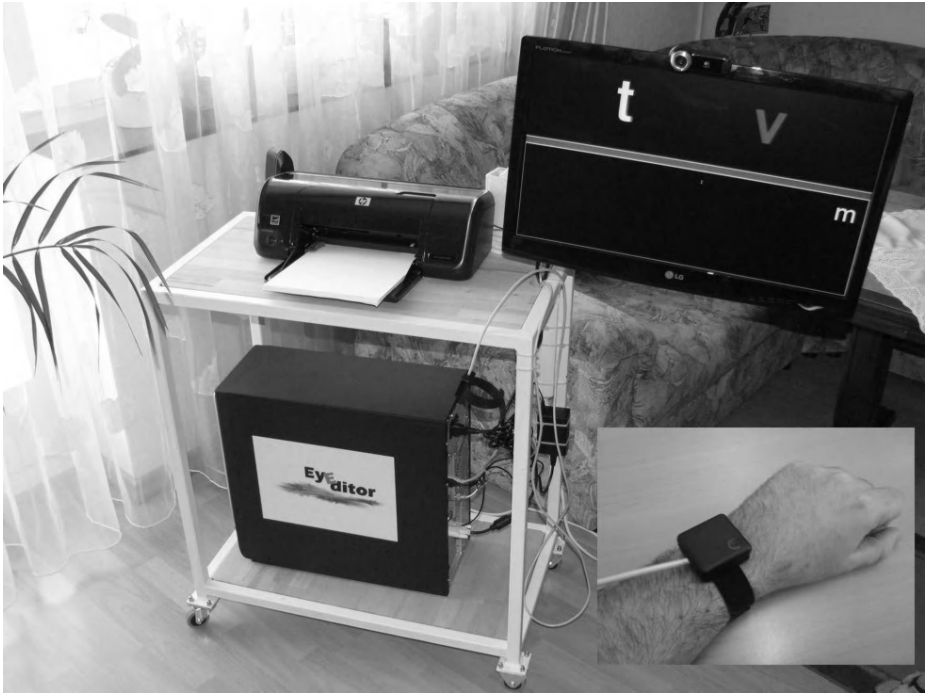


Fig. 6. Text writer system with camera, sensor and printer

lighting conditions and different computer hardware. By doing a series of experiments we found, that for people wearing glasses template matching for finding the pupils should be used, otherwise edge detection works better. If the surrounding is too dark, detection speed decreases and more errors arise. Normal room lighting is sufficient, some additional lighting would improve detection results. Concerning speed we can say, it is best to use a state-of-the-art personal computer (quad core), as slower computers didn't manage the calculations in realtime.

Using the acceleration sensor showed more reliable results, as in this case wrong detected signals nearly didn't occur. Also it seems to be preferable to use a motion sensors in combination with the blink detection, as blinking for selecting letters seems to be more stressful for the user than shaking a limb. The user becomes tired more quickly. So we propose to use signals generated by blinking for selecting words from line two, and the motion sensor for selecting letters from line one.

6 Conclusion

In our work we could demonstrate that our software is suitable for composing texts by eye blinking or motion detection. By using more than one input device,

writing speed could be increased. Our method is suitable for disabled people who can blink controlled like healthy ones or for people who are able to move some muscles. It is possible to extend our editor software by other sensors and input methods which generate binary input signals. Acoustic detectors may be considered. Also it is possible to change the user interface fitting special needs of the disabled individual. Mailing functions will be added in the near future, in order to give the possibility to get in touch with the surrounding world.

References

1. Morris, T., Blenkhorn, P., Zaidi, F.: Blink detection for real-time eye tracking. *Journal of Network and Computer Applications (JNCA)* 25, 129–143 (2002), doi:10.1006/jnca.2002
2. Chau, M., Betke, M.: Real Time Eye Tracking and Blink Detection with USB Cameras. In: Boston University Computer Science, Technical Report No. 2005-12 (2005)
3. Söllner, M., Hofmann, P., Pösl, J.: Using Triaxial Acceleration Sensors as Input Devices for Disabled Persons. In: ICCHP 2010, Computers Helping People with Special Needs Springer, Berlin (2010)
4. Schmidt, A., Kasinski, A.: The Performance of the Haar Cascade Classifiers Applied to the Face and Eyes Detection. In: *Computer Recognition Systems 2*, pp. 816–823. Springer, Heidelberg (2007), doi:10.1007/978-3-540-75175-5_101
5. Wilson, P.I., Fernandez, J.: Facial feature detection using Haar classifiers. *Journal of Computing Sciences in Colleges* 21(4), 127–133 (2006); ISSN:1937-4771

Augmented and Alternative Communication System Based on Dasher Application and an Accelerometer

Isabel Gómez, Pablo Anaya, Rafael Cabrera, Alberto J. Molina,
Octavio Rivera, and Manuel Merino

Electronic Technology Department, Universidad de Sevilla, Spain
{igomez,almolina}@us.es, {pjanaya,rcabreraac}@gmail.com,
{octavio,manmermon}@dte.us.es

Abstract. This paper describes a system composed by predictive text input software called “Dasher” and a hardware used to connect an accelerometer to the computer. The main goal of this work is to allow people with motor disabilities to have a flexible and cheap way to communicate. The accelerometer can be placed on any body part depending on user preferences. For this reason calibration functionality has been added to dasher software. The calibration process is easy and requires only some minutes but it is necessary in order to allow system can be used in different ways. Tests have been carried out by placing the accelerometer on the head. A rate of 26 words per minute is reached.

Keywords: dasher, accelerometer, open and flexible system, text input.

1 The R & D or Application Idea

The objective of this research project is to study dasher and its possibilities when it is handled by an accelerometer.

Dasher is a text input system which incorporate prediction¹ [1]. The system infers the most probable next characters, calculating the use frequency through a training text. During the program execution, letters move from the right to the center of the screen. Those letters are shown in color boxes which size depends on the probability of its occurrence (Figure 1). Dasher has different modes of operation: 2-D continuous input; 1-D continuous input; discrete input. This allows that different input devices could be used.

In 2-D mode, both values are used to move the cursor on the screen, and the speed depends on the relative position of the cursor respect to center. In 1-D mode, there is only one input, midrange values of the input control the zooming direction, values towards the extreme of the available range allow the user to zoom out and pause. In discrete input mode Dasher zoom in depends on which region the user select.

The accelerometer used for this application has two axis so 2-D mode of Dasher is possible. The system may be used by users with the ability to make small movements with any body part. Therefore, users with motor impairments who cannot use the standard keyboard and mouse are the main beneficiaries of this system.

¹ Dasher is free open source software.

complete dual-axis accelerometer with signal conditioned voltage outputs, which is all on a single monolithic IC. The product measures acceleration with a full-scale range of $\pm 2g$. The ADXL322 is an ultra small package (4 x 4 x 1.45 mm LFCSP).

Amplification and filtering board is connected to the accelerometer; the aim of this board is to eliminate quick and undesirable movements and it allows the system to work with a reduced angle range what increments system usability. The microcontroller-based system reads the accelerometer x-y coordinates and communicates that acceleration to a PC. Another function of this system is to perform digital signal processing. In order to increment the system effectiveness, we have programmed a moving average filter. Arduino has been used for implementing this system. Arduino is an open-source electronics prototyping platform based on flexible, easy-to-use hardware and software [12]. It is based on the Atmel Atmega328 microcontroller. The last element is the personal computer where dasher is installed.

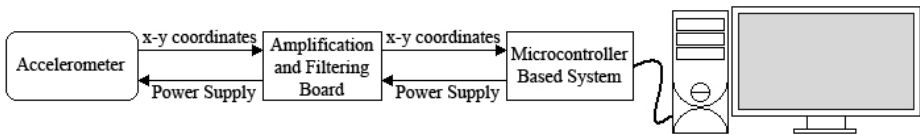


Fig. 2. System Architecture

3.2 Software Analysis and Design

Our aim in this work is the analysis of Dasher software to connect any kind of input device. The first step is to connect the accelerometer. In our very case, we are working in biosignal processing and we have as a main target the use of electromyography and electrooculography to control this text input software.

The first attempt was connecting dasher using socket as it was the most flexible option. Furthermore it is assumed that the software is ready for a configuration of this kind of input. However, the option above described is not implemented. Instead, we edited the dasher code to make it read an emulated serial port. In this way, the accelerometer reading is done via an USB port. Once the data has been obtained, it is necessary to adapt them to the values that Dasher uses to place the cursor on the screen: timing, scaling and adapting to screen resolution.

The accelerometer can be placed on anywhere depending on the user's preferences. For this purpose a calibration function was added to the Dasher code. (Fig.3).

The calibration process is easy and requires only a few minutes but it is necessary in order to allow the system to be used by different users. The user is asked to execute a circle movement in the widest way he is able to.

3.3 System Testing

The tests were carried out by placing the accelerometer on the head. In Figure 4 we can see the actual set up. We will be using WPM (word per minute) as our unit to measure the efficiency tests results.

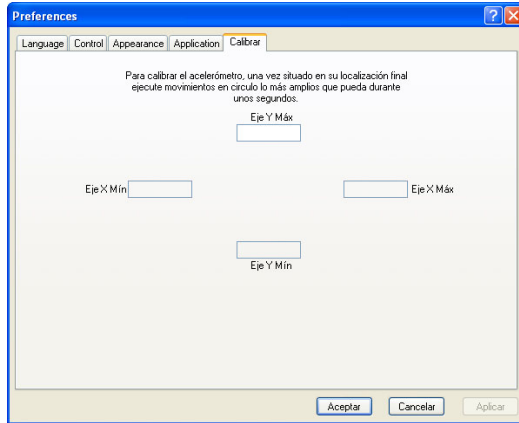


Fig. 3. Calibration Menu

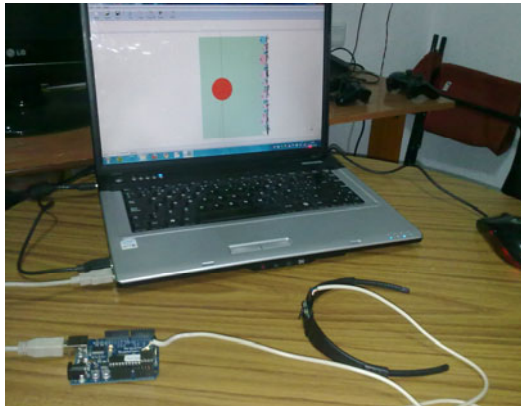


Fig. 4. Test System

Three non disable users (two experts and one novice) introduced a dictation text using the system; the system was trained with another text (the training test). Tests are performed in three cases depending on the characteristics of the training text: 1. without training text (untrained system) 2. text that doesn't match to dictation text, 3. text that is well matched to the dictation text.

A training text of 400 words and a dictation text of 40 words were used. A user is considered a novice when has not used the system previously, on the other hand an user reaches the expert grade with few uses of the system.

4 Results

The more the machine trains the more it “learns” and the users become more familiar with the usage of the application. Proficiency is achieved by increasing the usage time of the system with a text that the user is getting into. (See table 1).

Table 1. Results measured in WPM

Training cases	Expert1	Expert2	Novice
Untrained	2,6	3,4	2
Not matched	4,8	5,33	3,69
Matched	17,33	26	13

As we can see in [2], these are the results obtained when using Dasher with other types of input devices. For gaze use, these ranges go from 15 to 25 WPM, for driving with breathing 8 to 17 WPM and finally with the actual mouse which goes between 25 and 35 WPM. Therefore the results obtained with the accelerometer are within the range of those that can be achieved with other devices.

5 Conclusion and Planned Activities

In this paper it has presented the design and testing of a very low cost system that enables the communication to a disabled user. The system is uses a user-friendly interface and the user acquires a high level of expertise in a short period of time. Tests have been conducted with able-bodied users but we are working to test the system with disabled users and we hope to present the result of these tests in the final version of this application.

Two lines are opened as planned activities:

1. Connect Dasher with input devices based on biosignals.
2. Study the use of accelerometers in tele-rehabilitation systems design.

Acknowledgements. This project has been carried out within the framework of a research program: (p08-TIC-3631) – Multimodal Wireless interface funded by the Regional Government of Andalusia.

References

1. The Dasher Project, <http://www.inference.phy.cam.ac.uk/dasher>
2. Shorrocks, T.H., Mackay, D.J.C., Ball, C.J.: Efficient communication by breathing. In: Deterministic and Statistical Methods in Machine Learning, Springer, Heidelberg (2005)
3. Mackay, D.J.C., Ball, C.J., Donegan, M.: Efficient communications with one or two buttons. In: Fisher, R., Preuss, R., von Toussaint, U. (eds.) Proceedings of Maximum Entropy and Bayesian Methods, ser. Amer. Instit. Phys. Conf. Proc., vol. 735, pp. 207–218. Amer. Inst. Phys., Melville (2004)
4. Ward, D.J., Mackay, D.J.C.: Fast hands-free writing by gaze direction. *Nature* 418(6900), 838 (2002)
5. Wills, S.A., Mackay, D.J.C.: DASHER- An Efficient writing system for Brain-Computer Interfaces? *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 14(2) (June 2006)
6. National Center for Biotechnology Information, <http://www.ncbi.nlm.nih.gov>

7. Felton, E.A., Lewis, N.L., Wills, S.A., Radwin, R.G., Williams, J.C.: Neural Signal Based Control of the Dasher Writing System. In: Proceedings of the 3rd International IEEE EMBS Conference on Neural Engineering (2007)
8. Hamel, M., Fontaine, R., Boissy, P.: In-Home Telerehabilitation for Geriatric Patients. *IEEE Engineering in Medicine and Biology Magazine* 27(4) (July/August 2008)
9. Nakazawa, N., Yamada, K., Matsui, T., Itoh, I.: Development of Welfare Support-Equipment for Personal Computer Operation with Head Tilting and Breathing. In: IECON Thirty-First Annual Conference of the IEEE Industrial Electronics Society (2005)
10. Chen, Y.-L.: Application of Tilt Sensors in Human-Computer Mouse Interface for People with Disabilities. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 9(3) (September 2001)
11. Sad, H.H., Poirier, F.: Evaluation and modeling of user performance for pointing and scrolling tasks on handheld devices using tilt sensor. In: Second International Conferences on Advances in Computer-Human Interactions, IEEE Computer Society, Los Alamitos (2009)
12. Arduino, <http://www.arduino.cc/>

Evaluation of WordTree System with Motor Disabled Users

Georges Badr and Mathieu Raynal

IHCS / IRIT, Université Paul Sabatier, 31062 Toulouse cedex 9, France
{badr, raynal}@irit.fr

Abstract. In this paper we propose a novel interaction technique to use with a prediction list. Our aim is to optimize the use of this assistive aid. A prediction list is generally used to reduce effort by presenting a list of candidate words that may continue the prefix of the user. However, this technique presents a disadvantage: sometimes the word that the user is aiming to write is not presented in the list. The user has to continue his typing manually. Our technique allows the user to click on any letter in a word of the list so that the substring from this letter is inserted in the text. We also present the experiments carried out to compare our system to the classical list.

Keywords: Soft keyboard, interaction, word list prediction, motor disabled people.

1 Introduction

People with severe motor disabilities (like myopathy, cerebral palsy, musculoskeletal disorders, etc.), have difficulties using physical keyboards. Generally, they are obliged to use on-screen virtual keyboards. Two main types of interaction are possible: selecting characters (or others items) thanks to a pointing device such as mouse, trackball, eye tracking, etc. ; or using an automatic cursor which successively highlights each key. The user hits on a single switch input device when reaching the desired key. However, in both cases, text input is often slower than with a physical keyboard. To remedy this problem, many solutions have been tested since the early 80s. The most popular solution used is the word list near the keyboard and containing words most likely (cf. Fig. 1).

There are different kinds of prediction lists: those which are displayed in a fixed list as UKO-II [1], or close to the cursor in the soft keyboard, like PoBox [2] (cf. Fig. 2) or displayed in a popup which appears nearby the writing cursor [3]. Moreover, these lists could be used with pointing devices [2,4] or with a scanning system [5]. These lists are coupled to a word (or text) prediction system. These systems are also the subject of many researches – in particular in speech processing [6,7].

However, whatever the prediction system used, all user interfaces connected to a word prediction system allow the user to select a whole word, but not a prefix of a word. The user has to complete the text entry until the aimed word is displayed in the list. The present paper focuses on this problem of interaction in the word list: the

selection of prefix. To remedy this shortcoming, we have implemented a technique (called WordTree) which allows the user to select a part of the word in the list, and to insert the selected substring in the text.

This paper introduces a technique that allows the user to select a substring of a word displayed in a list with a pointing device. After the description of the principle and of the architecture of this system, we will present the results of the experiments.

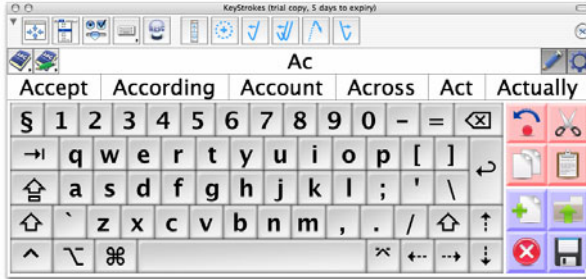


Fig. 1. Keystrokes [4]



Fig. 2. Prediction list of the PoBox [2]

2 WordTree

2.1 Principle

Consider the situation where the aimed word is not shown in the list. The user has then to continue his writing, waiting for his word in the list, and sometimes the user has to complete it manually. Our idea is to conceive a new interaction technique with the prediction list to help the user find his word faster. This technique, called WordTree, allows the user to select a substring of the word, without being forced to select the whole word. Once any other word is listed and contains part of his word, he can select this substring so he can accelerate his typing.

Each character in the listed word is represented by a clickable node (cf. Fig. 3). When a node is clicked, the substring containing all the nodes between the root and the selected node is inserted into the text. For example, if the user wants to write the

word "complex", he starts typing "co" on the keyboard. A list of candidate words is displayed. For some reason, (i.e. a low probability of this word in the prediction system), suppose that the word "complex" is not shown in the list. Instead we have the word "complete". In order to advance in his text, the user can click on the first 'e' of the word "complete" (cf. Fig. 3) and the string "mple" is inserted in the text. In this way, the user can reduce the number of operations and consequently increase his text input speed.

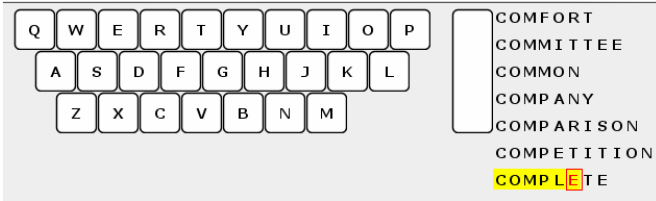


Fig. 3. WordTree interface

2.2 Architecture

WordTree is divided into three modules: the soft keyboard, the prediction system and the presentation of predicted words. Our contribution is on the presentation interface. This modular architecture allows to easily adapt our system to any type of soft keyboard. Its advantage comes from the fact that it is independent of the prediction algorithm: it can be used with any prediction system.

2.3 Prediction System Used

In this study, WordTree is allied to a prediction system based on a lexicographic tree constructed from a set of words (dictionary). The tree is composed of 364,370 words. Each word is represented by a path from the root to a leaf of the tree, in which each

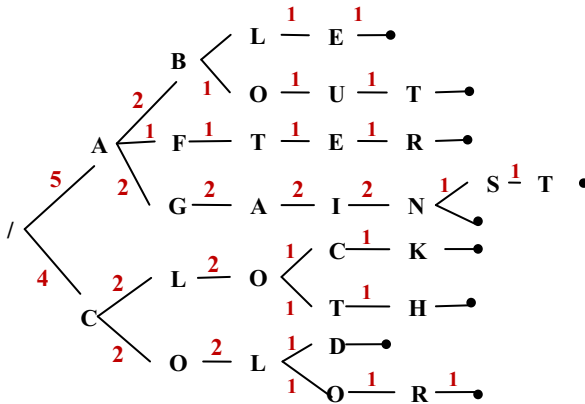


Fig. 4. Lexicographical tree with words: able, about, after, again, against, clock, cloth, cold, colour

character is a node of the tree (cf. Fig. 4 **Fehler! Verweisquelle konnte nicht gefunden werden.**). Two words with the same prefix use the same initial path to avoid creating too many new nodes. For words which are prefixes for other words, a special leaf “end word” was created (black circle on the Fig. 4). Each arc of the tree is weighted: when a word is added, the system inserts each character in its place by crossing the tree according to what was already entered. The arcs between nodes which are crossed are incremented by 1.

When the user enters a prefix, the system classifies the characters which can succeed according to their probability of appearance (the biggest the value of the arc is, the more the character has chance to appear). This weight will let it sort the predicted words in the list according to their probability defined by their weights.

3 Method

We hypothesized that WordTree would allow reducing the number of clicks necessary for the text entry. Moreover, if the user has less operation to be performed, we believe it will accelerate its text input speed. A first simulation shown that this technique can theoretically reduce the number of clicks by 25% [8].

This study was undertaken to evaluate WordTree. Since QWERTY is the most common layout, we included both layouts in the study.

3.1 Participants

Eight male volunteers with severe motor disabilities participated in this experiment. They were all regular users of desktop computers with QWERTY soft keyboards.

3.2 Apparatus

The experiment software was developed with Java SE 6. Each subject used his own PC and pointing devices to interact with the soft keyboard. Three of them used a classical mouse and the five others pointed with a trackball.

3.3 Design

Each session consists of two exercises of copy on a QWERTY keyboard only with the 26 characters of the Latin alphabet and the space bar (cf. Fig. 6): one of the two with the classical system (cf. Fig. 5 on left) and the second one with WordTree (cf. Fig. 5 on right). WordTree, as well as the classical list, are using the same prediction algorithm described above. There was a training session for each exercise before the first session.

Subjects had to copy out 30 of the most usually used words (215 characters), which were the same for the two exercises and each session. Subjects were instructed to proceed as quickly and as accurately as possible. Each subject attended 10 sessions over a period of 8 to 10 days. Sessions were spaced by no less than three hours and by no more than two days. We used a counterbalanced, within-subjects design.

The independent variable was the type of word prediction list, either classical list or WordTree. Dependent variables were text entry speed (character per second - CPS), keystroke per character (KSPC) and error rate.

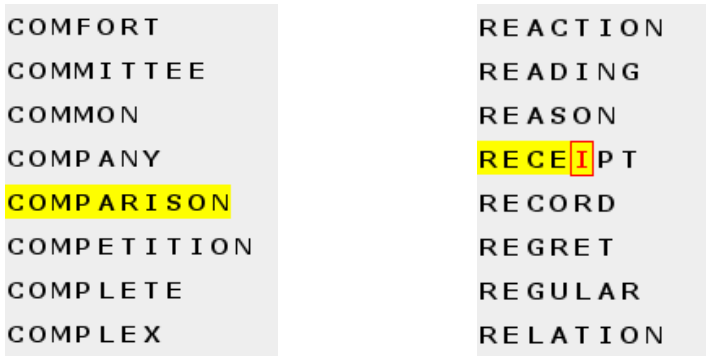


Fig. 5. different lists studied: classical list on left and WordTree on right

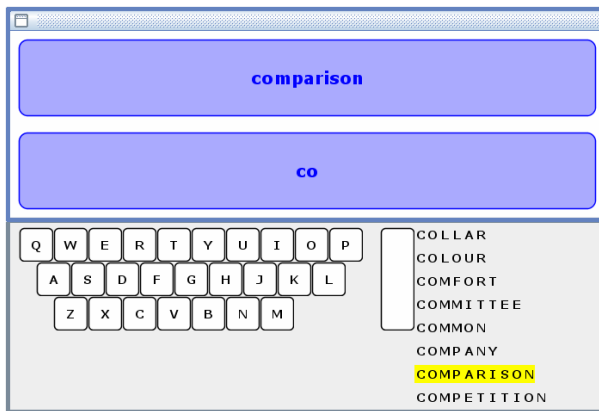


Fig. 6. Experiment screen with classical list

3.4 Procedure

The word to be copied was presented on a line, and the word being typed by the user appeared on the line below. The text entry errors were not displayed on the screen. Instead there was a visual and audio feedback signaling the error and the strip did not move until the subject entered the right character. At the end of each word, the participant had to hit the space bar.

4 Results and Discussion

The statistical results are presented in this section and were analysed using the Analysis Of VAriance (ANOVA) test.

4.1 Text Entry Speed

The analysis of those results (cf. Table 1) shows that our hypothesis was confirmed. In fact, we can see a reduction of the number of operations needed to be done with

WordTree compared with the classical list. On average, KSPC measure is equal to 0.825 with the classical list, whereas it is only 0.663 with WordTree. Thus WordTree reduces the number of keystrokes on average by 19,6% ($F(1, 158)=1334,4$, $p<0,0001$).

Moreover, this reduction of keystrokes causes an improvement of the text input speed with WordTree. On average, participants entered 0,566 characters per second with the classical list, whereas they entered 0,604 characters per second with WordTree. Thus they gain on average 6.7% of time with WordTree compared with classical list ($F(1, 158)=7.7142$, $p=0.00614$). The Fig. 7 shows that sessions have no impact on the difference of performance between classical list and WordTree. Both technique increase equally. There is no significant technique-by-session interaction ($F(9, 140)=,17419$, $p=,99635$).

Table 1. Performance with the different techniques

	KSPC	CPS
Classical list	0.825	0.566
WordTree	0.663	0.604

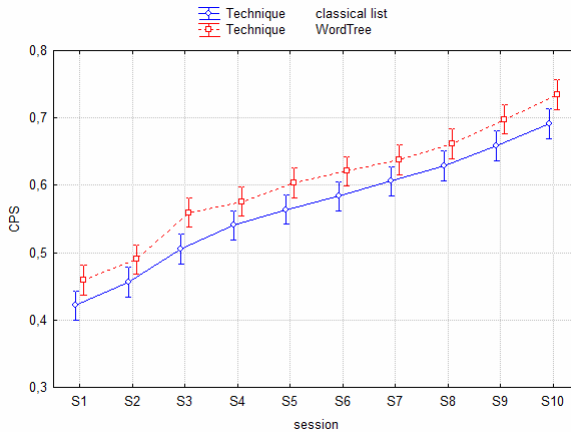


Fig. 7. Text entry speed by technique and session

To conclude in text entry speed, Fig. 8 shows that device used by participants has no impact on the text entry speed ($F(1, 156)=0.00715$, $p=0.93272$).

4.2 Error Rate

During the exercises, when the current character differed from the expected character, the error was recorded. The average of error rate was 2,42% for the classical list and 2% with the WordTree system ($F(3, 138)=670.51$, and $p<0.0001$).

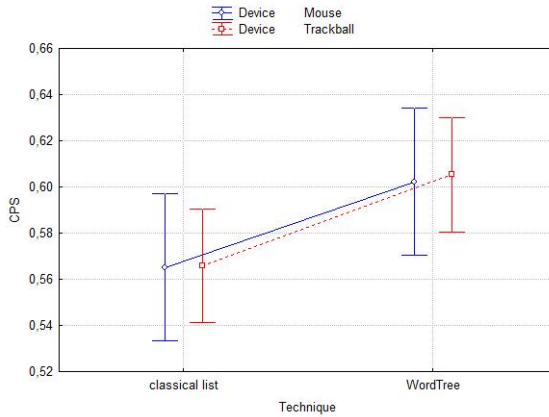


Fig. 8. Text input speed by techniques and devices

5 Conclusion

In this article, we presented a new interaction technique to be used in a list of predicted words. Unlike the classical list where the user is obliged to select an entire word in the list, the user can click on any character of a word in the WordTree list. The system selects the substring of this word from its root to the clicked node or character. The selected substring is then inserted into the text.

We also presented the results of the evaluation conducted with 8 disabled people. The results reported by our experiments show that the WordTree system can be beneficial and advantageous for text entry compared to the classical list of prediction. In fact, we have obtained a significant text entry speed augmentation and an important reduction of the number of operations necessary for entering a word. Thus, with WordTree, a given word could be entered with a less number of interactions, and less moving through the keystrokes of the keyboard. This task was so tiring for both disabled and non-disabled people.

References

1. Harbusch, K., Hasan, S., Hoffmann, H., Kühn, M., Schüler, B.: Domain-specific disambiguation for typing with ambiguous keyboards. In: Proceedings of the 2003 EACL Workshop on Language Modeling For Text Entry Methods. ACL Workshops, pp. 67–74. Association for Computational Linguistics, Morristown (2003)
2. Masui, T.: An efficient text input method for pen-based computers. In: Proceedings of the ACM Conference on Human Factors in Computing Systems CHI (1998)
3. Beck, C., Seisenbacher, G., Edlemayer, G., Zagler, W.L.: First User Test Results with the Predictive Typing System FASTY
4. Keystrokes on-screen keyboard, user manual, <http://www.assistiveware.com/files/KeyStrokesXmanualA4.pdf>

5. Wandmacher, T., Antoine, J.-Y., Poirier, F., Departe, J.-P.: SIBYLLE, An Assistive Communication System Adapting to the Context and Its User. *ACM Transactions on Accessible Computing* 1(1), 1–30 (2008)
6. Aliprandi, C., Carmignani, N., Mancarella, P.: An inflected-Sensitive Letter and Word Prediction System. *International Journal of Computing & Information Sciences* 5(2), 79–85 (2007)
7. Hunnicutt, S., Nozadze, L., Chikoidze, G.: Russian Word Prediction with Morphological Support. In: 5th International Symposium on Language, Logic and Computation, Tbilisi, Georgia (2003)
8. Badr, G., Raynal, M.: WordTree: Results of a Word Prediction System Presented thanks to a Tree. In: International Conference on Universal Access in Human-Computer Interaction, San Diego. LNCS, vol. 5616, pp. 463–471. Springer, Heidelberg (2009)

Evaluation of SpreadKey System with Motor Impaired Users

Bruno Merlin and Mathieu Raynal

IRIT, Université Paul Sabatier,
118 Route de Narbonne, F-31062
TOULOUSE CEDEX 9
{raynal,merlin}@irit.fr

Abstract. We recently presented a new soft keyboard: SpreadKey. The keyboard is based on a QWERTY layout and dynamically recycles the needless characters. The needless characters are determined by a predictive system in function of the previous input characters. The recycling aims at providing several times the characters with a high probability to be typed. This article resumes the concept of SpreadKey and explains the experimentation we conduct to evaluate it. The experimental results show that SpreadKey is an interesting solution to assist motor impaired users (with motor impairment of the superior limbs) in the text input task.

Keywords: Text entry, Fitts' law, soft keyboard, assistive technology.

1 Introduction

Many motor impaired users of the superior limbs are not able to use a standard physical keyboard. For the more severe impairment, the last issue to input text is using single switch keyboards such as Sibylle [1] or FOCL [2]. But many of the other users are able to interact with a soft keyboard through a mouse or a trackball. However, according to the Fitts' law [3], several researches [4,5] demonstrated that the implementation of the QWERTY keyboard as soft keyboard is not an optimized way to input text through a pointing device. They proposed alternative layout [4,5] aiming at reducing the distances covered by the cursor or ambiguous keyboards [6,7] aiming at both reducing the distance and increasing the keys.

But, even if they potentially improved by 40% [4] the performances of an expert user, we observed that these new layouts are not easily adopted. Two main reasons explain it. On the one hand, users are accustomed to use the QWERTY for a long time and switching to a new soft keyboard involves a learning cost. During the learning period, the user must overpass a frustrating inefficiency. On the other hand, the QWERTY keyboard is the cultural reference implemented in every system. Consequently, the user of an alternative soft keyboard remains in contact with the QWERTY when he leaves his personal ambient.

Based on these observations, we recently proposed and presented a new soft keyboard: SpreadKey [8]. The keyboard displays a QWERTY layout dynamically

optimized according to the input context. After each keystroke, the keyboard replaces the characters with a very low probability to be input by a character with a high probability of occurrence. The algorithm aims at increasing the key size for characters with high probability of occurrence by aggregating several keys displaying these characters and to reduce the distance covered by the cursor. But it keeps the QWERTY structure as reference.

In a first step [8], we evaluated the theoretical keyboard performances toward the Fitts' law [3]. We calculated the lower and upper-bound [9]. However, this theoretical evaluation neglects cognitive impact due to the dynamic changes. We consequently initiated an experimental evaluation of the keyboard. This article presents the results of this evaluation.

In this article, we resume the concepts of SpreadKey and the conclusions of the theoretical evaluation. Then, we present our experimental protocol and, finally, we describe and analyze the results of this experimentation.

2 SpreadKey Concepts

SpreadKey targets two principal objectives: At first, it tends to reduce the distances covered by the cursor. Like KeyGlass [10], SpreadKey aims at providing new keys displaying the most probable character near the cursor. At second, like FishEye Keyboard [11], SpreadKey tends to increase the key sizes in order to ease the pointing.

The keyboard is based on a standard layout (QWERTY) and uses a prediction system to classify the characters by probability of occurrence. The characters with a very low probability (closed to 0) to be typed are replaced by characters with high probability to be typed. The recycling algorithm tends to:

- Insert new occurrences of the high probable characters between the last pressed key and the key containing normally the high probable character (cf. figure 1). The end is to reduce the distance covered by the cursor.
- Aggregate new occurrences of the high probable characters in the neighborhood of the key containing normally the high probable character (cf. figure 2). The aim is to increase the size of the targets.

Thus, in several situations, mainly for the input of the end of words, the whole keyboard is recycled in one or two large keys. In these contexts, the probability for the last pressed key being recycled with the next required key becomes very high. It decreases considerably the distance covered by the pointing device.

Nevertheless, in order to prevent from prediction errors (words from a foreign language, names or acronyms etc.), SpreadKey provides two alternative interactions (a long-press or a flick in direction of the top-left corner of the key) to override the prediction and to access to a replaced character. The substitution of an improbable character aims at creating an aggregation of keys providing the same character (equivalent to a key with a higher size).

Consequently, after each stroke, the keyboard layout is dynamically adapted according to the new context (cf. Figure 3). However, each character remains accessible at its own place enabling the user to anticipate the next stroke. The characters are

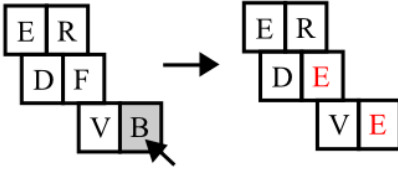


Fig. 1. The recycling algorithm tends to insert new occurrences of characters with a high probability to be typed between the next key pressed and the key that normally contains this character

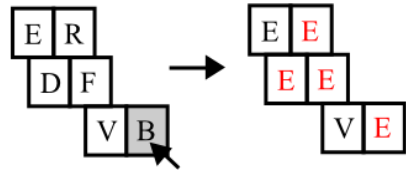


Fig. 2. The recycling algorithm tends to aggregate new occurrences of the high probable characters in the neighborhood of the keys that normally contains the characters

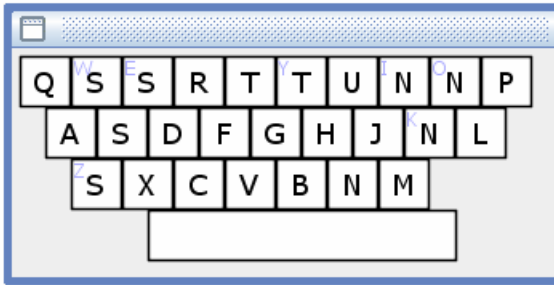


Fig. 3. SpreadKey is based on a QWERTY keyboard. But, after the input of the prefix “E”, the characters W, E, Z, Y, I, O and K are improbable in English. So, they have been replaced by the characters S, T or N that get the higher probability of occurrence. Nevertheless, the original characters remain accessible by a long-press or a flick.

accessible through standard, ambiguous keys (for character with a very low probability to be input and is remapped), or aggregations of keys (if the character get a high probability to be input).

In a first step, we compared the theoretical performances of SpreadKey with the theoretical performance of the QWERTY soft keyboard [8]. We calculated the theoretical performances of QWERTY according to the Soukoreff and Mackenzie methodology [9]. Because of the dynamicity of the SpreadKey layout, we could not use the same calculus to obtain the SpreadKey’s performances.

To establish the SpreadKey’s theoretical performances, we simulated the input of a large set of texts. During the simulation, we measured the text input speed according to the Fitts’ law [3] (adapted by Mackenzie [12] for HMI) and measured the distance theoretically covered by the cursor.

We calculated the two values with different parameters for the Fitts’ law. These different parameters emulated different levels of motor impairment.

The results were very encouraging (cf. table 1). In comparison with QWERTY, SpreadKey should reduce the distance covered by the cursor (about 45%) and reduce significantly the input time (between 35% and 50%). The simulation with different parameters for the Fitts’ law showed that the ratio should increase according to the user impairment. The more the motor impairment is severe, the more SpreadKey should be efficient.

Table 1. Theoretical input speed results. The bandwidth characterizes the handicap impairment. 5 correspond to a normal user, 0.5 correspond to very severe motor impairment.

Bandwidth	0.5	1	2	3	4	5
SpreadKey/Qwerty input time (%)	50.9	52.2	54.8	57.9	61.2	64.5
WPM SpreadKey	5.5	10.8	20.5	29.2	36.7	43.5

However, the model used to calculate the input time does not take into account cognitive perturbations due to dynamic changes. So the results must be confirmed by experimental evaluations.

3 Experimentation

The aim of the experimentation was to evaluate the performances of SpreadKey with motor impaired users. According to the result of the simulation and pre-experimentation, we hypothesized that SpreadKey would allow reducing the distance covered by the pointer, increasing the key size and improve the input speed.

3.1 Participants and Apparatus

Six male volunteers with severe motor impairment of the superior limbs participated to this experiment. They are all regular users of desktop computers with QWERTY soft keyboards and interact with computer through a trackball.

The experiment environment is the new version of the plate-form E-Assist [13] developed in Java. The platform proposes an XML language to model the keyboards behavior and a set of components to create experimentations. The keyboards are rendered with the library IntNovate [14]. The schedule of the experimentation: sessions, order of exercises, user groups are equally modeled as an XML scenario.

Each subject used a desktop computer and a trackball to interact with the soft keyboards during the experiment.

We created manually a sentence set to be entered. The sentence set contains 30 sentences in the user's native language. The sentences respect the character and digram frequencies of the language with a high level of correlation (97% for the character frequency and 91% for the digram).

3.2 Design

Each subject attended 10 sessions over a period of 5 days (two sessions per day). Sessions were spaced by no less than three hours. Each session consists in two exercises of copy: an exercise with a QWERTY soft keyboard reduced to the 26 Latin alphabetic keys and the space, and the second one with SpreadKey. There was a training session for each exercise before the first session.

The users were divided in two groups. A group began the first session with QWERTY, and the other group by SpreadKey. The users alternated the order of the exercises between each session.

Subjects had to copy out the sentences contained in the sentence set. They were the same for the two exercises and for every session. Subjects were instructed to proceed

as quickly and as accurately as possible. The independent variable was the type of keyboard, either QWERTY or SpreadKey. Dependent variables were text entry speed in character per second (CPS), distance covered by the pointer and error rate.

The sentence to be copied was presented in a strip (cf. figure 4), and the sentence being typed by the user appeared in a strip below. The text entry errors were not displayed on the screen. Instead there was a visual and audio feedback signaling the error and the strip did not move until the subject entered the right character. At the end of each sentence, the participant had to hit the space bar.

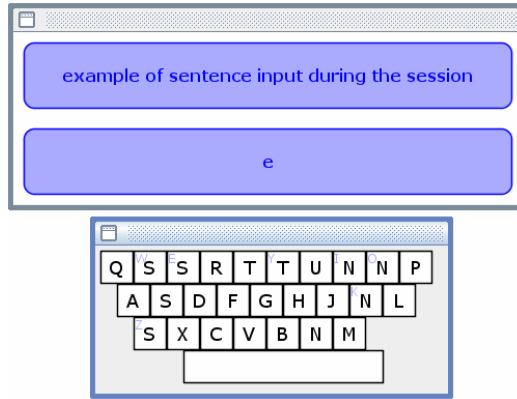


Fig. 4. Experimentation platform

4 Results and Discussion

The statistical results (cf. figure 5) are presented in this section and were analysed using the ANalysis Of VAriance (ANOVA) test. Data show that the order in which the exercises were performed had no impact on the results.

4.1 Distance

On average, the distance is equal to 213 pixels with the QWERTY keyboard, whereas it is only 138 pixels with SpreadKey (cf. figure 6). Thus SpreadKey reduces the distance on average by 35% ($F(1, 2398)=3227.3, p<0.0001$). The distance is 8.5% superior to the distance expected in the simulation with QWERTY and 3.6% superior with SpreadKey. Consequently, the benefice is proportionally superior to the benefice expected during the simulation (about 35% against about 30%). It illustrates that SpreadKey reduces the distance between the targets and, by increasing the target size, we observed that it diminishes the overshoot of the target during the pointing.

4.2 Text Entry Speed

On average (cf. figure 7), participants entered 0.87 characters per second (10.44 wpm) with the QWERTY keyboard, whereas they entered 1.09 characters per second with

SpreadKey (13.08 wpm). Thus they gain on average 20% of time with SpreadKey compared with QWERTY keyboard ($F(1, 2397)=723.92, p<0.001$).

Even if, the input speed with SpreadKey is higher than the input speed obtained with QWERTY, this value remains significantly lower than the value expected toward the simulation. Two main reasons explain these results. At first even if the user can anticipate a move toward the next stroke because every character remain accessible at its own place, the users confirmed that the dynamic changes perturb this anticipation. At second, the users explained that, many times, they preferred not to anticipate the move toward the next stroke and targeted the nearest key. They explained that this strategy was less efficient (considering the input speed) but more comfortable.

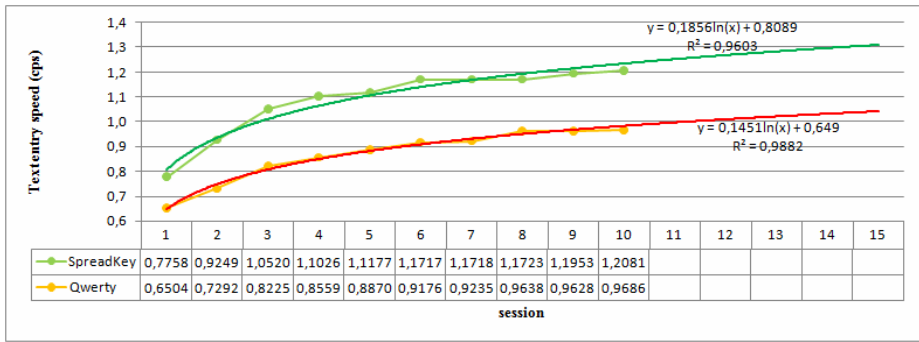


Fig. 5. Results for each session

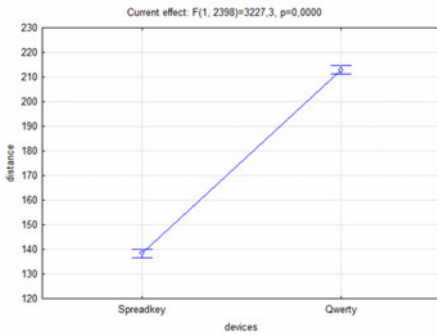


Fig. 6. Averages for distance

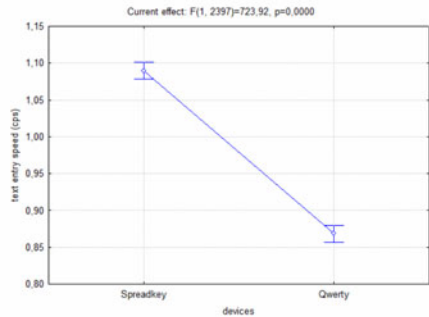


Fig. 7. Averages for input speed

The users mentioned too that the bounds of the aggregation of keys displaying the same character were not easily identified and expressed the necessity to identify graphically a group of keys displaying the same character.

4.3 Error Rate

Because SpreadKey increases the target sizes and reduces the distances covered by the pointing device, SpreadKey ease the pointing and we expected a lower error rate

with SpreadKey than with QWERTY. The results contradicted this hypothesis. The average of error rate was 1.5% for the QWERTY keyboard and 2% with SpreadKey ($F(1, 2398)=720.92$, and $p<0.0001$).

The users' comments, confirmed by a discrete analysis of errors, shows that a lot of errors are due to the users' anticipation of the next stroke toward the remapping algorithm. We identified the two main scenarios generating errors.

The first scenario is the input of two identical characters consecutively: a user must input two 'R'; in the current input context the 'E' is remapped as 'R'; to input the 'R', the user presses the 'E' remapped as 'R' and correctly input the first 'R'; in the new current input context, the 'E' is not remapped as 'R' anymore but the user anticipated the second click and presses the key 'E' a second time generating an error.

The second scenario is the input of some long words termination: for instance, the user wants to input the word "anticipation"; he inputs the prefix 'antici'; the only possibility in English is now to input a 'P', so, the whole keyboard is now remapped with "P"; the user pressed any key to input 'P'; now the only possibility is to input a 'A'; the whole is remapped with 'A', then with 'T'; But after the input of 'anticipat', different alternatives appear: 'anticipatOry', 'anticipatiOn', 'anticipatE'; Encouraged by the three previous inputs, the user tends to anticipate the next click before understanding that the key is not necessarily remapped with the good character anymore.

5 Conclusion and Future Works

After a theoretical evaluation of SpreadKey, we proceed to an experimental evaluation involving users, motor impaired of the superior limbs.

Even if the performances were lower than the performances expected toward the input time, the experimentation confirmed the good capabilities of SpreadKey to help our users in the text input task. By increasing the key size and reducing the distance covered by the cursor, SpreadKey turns the input easier and less tiring physically.

However, the experimentation highlighted two main problems increasing the cognitive charge of the user. At first, the remapping algorithm induces users' errors in some identified contexts (cf. previous section). At second, the graphical representations of aggregation of key displaying the same character do not ease the analysis of changes operated during the remapping.

Our future works will lead us to solve these problems by adding visual clues.

References

1. Wandmacher, T., Antoine, J., Poirier, F., Départe, J.: Sibylle, An Assistive Communication System Adapting to the Context and Its User. *ACM Trans. Access. Comput.* 1(1), 1–30 (2008)
2. MacKenzie, I.S.: Mobile text entry using three keys. In: *Proc. of Nordi CHI 2002*, pp. 27–34 (2002)
3. Fitts, P.M.: The information capacity of the human motor system in controlling the amplitude of movement. *Journal of Experimental Psychology* 47, 381–391

4. Raynal, M., Vigouroux, N.: Genetic algorithm to generate optimized soft keyboard. In: CHI 2005 extended abstracts on Human factors in computing systems, Portland, OR, USA, April 2-7 (2005)
5. Zhai, S., Hunter, M., Smith, B.A.: The metropolis keyboard - an exploration of quantitative techniques for virtual keyboard design. In: Proceedings of the 13th annual ACM symposium on User interface software and technology, San Diego, California, United States, November 6-8, pp. 119–128 (2000)
6. Levine, S.H., Goodenought-Trepagnier, C., Getschow, C.O., Minneman, S.L.: Multi-character key text entry using computer disambiguation. In: Proceedings of the 10th Annual Conference of Rehabilitation Engineering, pp. 177–179. RESNA, Washington
7. Harbusch, K., Hasan, S., Hoffmann, H., Kuhn, M., Shuler, B.: Domain specific Disambiguation for Typing with Ambiguous Keyboards. In: EACL 2003 Workshop on Language Modeling for Text Entry Methods, Budapest, April 14 (2003)
8. Merlin, B., Raynal, M.: SpreadKey: increasing software keyboard key by recycling needless ones. In: AAATE 2009, Florence, Italia (2009)
9. Soukoreff, R.W., MacKenzie, I.S.: Theoretical upper and lower bounds on typing speeds using a stylus and soft keyboard. *Behaviour & Information Technology* 14, 370–379 (1995)
10. Raynal, M., Vigouroux, N.: KeyGlasses: Semi-transparent keys to optimize text input on virtual keyboard. In: AAATE 2005, Lille, September 6-9, pp. 713–717 (2005)
11. Raynal, M., Vinot, J.L., Truillet, P.: Fisheye Keyboard: Whole Keyboard Displayed on Small Device. In: UIST 2007, Newport, USA (2007)
12. Mackenzie, I.S.: Fitts's law as a research and design tool in human-computer interaction. *Human-Computer Interaction* 7, 91–139
13. Raynal, M., Maubert, S., Vigouroux, N., Vella, F., Magnien, L.: E-Assiste: A platform Allowing Evaluation of Text Input System. In: 3rd Int. Conf. on Universal Access in Human-Computer Interaction (UAHCI 2005), Las Vegas, USA, July 22-27, Lawrence Erlbaum Associates (LEA), Mahwah (2005)
14. <http://www.intnovate.org>

An Integrated Text Entry System Design for Severe Physical Disabilities

Yun-Lung Lin¹, Ming-Chung Chen², Yao-Ming Yeh¹, and Chih-Ching Yeh¹

¹ Graduate Institute of Information and Computer Education,
National Taiwan Normal University,
No.162, Sec. 1, Heping E. Rd., Da-an District, Taipei City 106, Taiwan
{harrison, ymyeh, ccyehe}@ice.ntnu.edu.tw

² Department of Special Education, National Chiayi University,
300 University Road, Chiayi, Taiwan
mtchen@mail.ncyu.edu.tw

Abstract. Text entry is an important and prerequisite skill for people to utilize the computer information technology; however, people with severe physical disabilities need appropriated text entry system to interact with computer. The authors designed a text entry system which integrated on-screen keyboarding and encoding technique. Furthermore, this integrated text entry system provided learning strategy and on-screen keyboard layout adaption tools. This study also introduced the effect of the application of this text entry system on two clients with severe physical disabilities. The results demonstrated that the text entry system benefited the clients.

Keywords: text entry, severe physical disabilities, on-screen keyboard, encoding technique.

1 Introduction

Text entry is essential for people to utilize the computer information technology in performing their daily routine. However, for people with severe physical disabilities, their physical impairments make text entry difficult and thus obstruct their use of computer technology.

Various kinds of alternative text entry devices and software are available for the people with severe physical disabilities. On-screen keyboard and encoding system are two popular text entry systems. Both commercial and research-based on-screen keyboards provide multiple and flexible access solutions for user with various restrictions of motor control [1]. Encoding system is also regarded as an efficient text entry for those with better motor control capability [2].

Though assistive devices and software ease the difficulties that physically-impaired users experience in interacting with computers, constraints abound with previous systems. Based on the authors' analysis, there are three limitations at least. First, though most on-screen keyboards offer flexible layout settings, none of them provide assessment tool to help detect the proper size of layout for the client. Thus, service provider is required to do the detecting manually and unsystematically. Second, since

no encoding system embeds learning strategy to assist the user in using the device, extensive practice is required of novice users to become familiar with the mapping codes of each character before they can generate the text. Third, no system has considered the dynamic needs of a client. For instance, current alternative text entry systems do not integrate on-screen keyboard and encoding system. Therefore, a client with degenerating motor control capability who no longer could use encoding system would have no choice but to learn a new on-screen keyboarding system later. Given the important approach of “Design for Dynamic Diversity” [3], the alternative text entry system should also attend to the client’s individual needs. Besides, the text entry system needs to embed assessment and learning strategy so as to facilitate the clinical professionals in their work of providing proper environment to the clients. To address the above issues, the current study aims to develop an innovative text entry system to remedy the constraints and inadequacy of the popular on-screen keyboards and encoding systems mentioned above.

2 The Text Entry System

This system adopted the idea of “Design for Dynamic Diversity”. Aside from the popular multiple selection methods (pointing-and-click, dwell, scanning), resizable layout, and flexible layout arrangement, this system integrated on-screen keyboard and encoding system into a single but flexible text entry system: an on-screen keyboard software as well as a numeric encoding text entry system when plugged in a numeric-based keypad or switches. To determine the proper size of the on-screen layout, the system is equipped with a layout adaptation evaluation program. Furthermore, the system employs visual prompt strategy to help the novice users to use numeric encoding system without spending lengthy time learning the rules of numeric combination.

2.1 The Layout Adaptation Evaluation Program

Layout size is very important when considering movement speed and accuracy simultaneously, as illustrated by Fitts’ Law [4]. The text entry system incorporated a layout adaptation testing program based on the binary search tree algorithm to determine the appropriate layout. The user could choose proper layout size from seven different layout sizes: 740^2 , 650^2 , 560^2 , 470^2 , 380^2 , 290^2 , and 200^2 pixels in a $1024 * 768$ resolution. A layout with a middle size ($470*470$ pixels) was given in the initial test. A smaller layout was presented when the initial test result reached the criterion of 90% accuracy. On the contrary, a larger layout was offered when the user cannot reach this criterion.

The layout of adaptation evaluation program was shown in Figure 1. The purpose of testing in each layout was to detect: 1) the reachable range [*pressing keys sequentially*: x, (, I, b, (, x, I, (, b, I, x, b, x], 2) the accuracy of clicking the single key [*pressing keys sequentially* : n, o, k, l, o, n, k, o, l, k, n, l, n], and the adjacent keys [*pressing keys sequentially* : g, k, 5, 9, k, g, 5, k, 9, 5, g, 9, g]. The procedure was repeated until an appropriate layout size for the user was found. On average, the appropriate layout for the user could be found after three or four rounds of test.

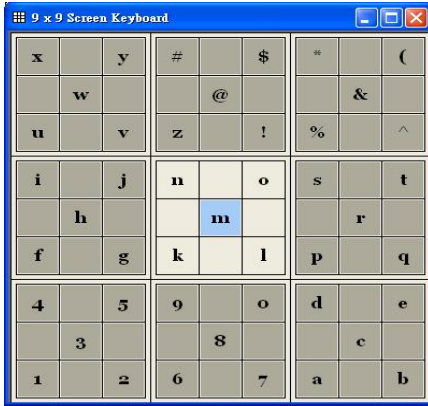


Fig. 1. The snapshot of the layout adaptation evaluation program



Fig. 2. The layout of the on-screen keyboard with alphabetical layout

2.2 The Encoding Text Entry System

In order to ease the novice user of his/her learning load , each character was represented by two Arabic numerals to make the mapping rule consistent in the encoding text entry system (ETE system). As the alphabetic layout shown in Figure 2, each character was generated by two numerals based-on the numeric keypad. For instance, the letter ‘a’ was generated when the number ‘4’ and ‘7’ are pressed in order. The ‘0’ was used as a cancel key when a typing error occurred.

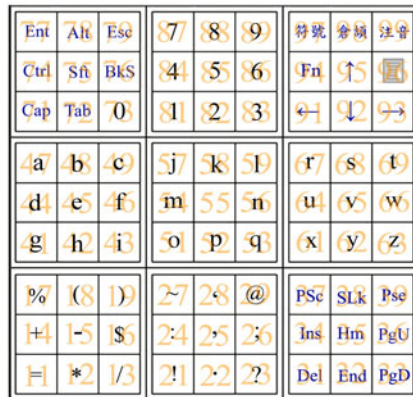


Fig. 3. The visual prompt layout of the encoding text entry system

Meanwhile, the visual prompt strategy provided an additional support. As Figure 3 showed, the visual prompts provided navigation mapping codes on the layout when the system was activated. The cues of number codes liberated the users from having to learn or memorize the mapping rule while training them to naturally and intuitively follow the mapping codes with practice. Moreover, such a design also served as a way

to notify the user either that he/she can proceed to press the next key or that he/she should modify the first key pressed wrong. In conclusion, the keyboard layout provided immediate visual feedback to the user's key pressing by highlighting specific areas or target keys.

3 Applications

Two severely physically-impaired adolescents in need of a more effective text entry system experimented on the use of this integrated system. Both obtained individualized support and performed effectively.

3.1 Application on Client with Osteogenesis Imperfecta

Amy was a female with osteogenesis imperfecta. Owing to her weak hand coordination, she cannot operate conventional keyboards well. She used the point-and-click selection method with her left hand; however, the text entry performance was poor. Though having three years computer experience for the purposes of leisure, she was not familiar with the QWERTY layout yet.

The experiment started with conducting layout adaptation evaluation program to detect a proper layout for her. The result indicated that level 3 (470* 470 pixels in a 1024 * 768 resolution) was the proper layout size. Her performance was 4.8 characters per minute (CPM) originally. Amy generated 5.6 CPM at the first trial by using the integrated text entry system. The performance sped up to 20.6 CPM with about 2 % error rate after the fifteen trials. For more detail about the research, interested readers may refer to our previous study [5].

3.2 Application on Client with Muscular Dystrophy

3.2.1 Participant's Profile

A 12-year-old sixth grader at a special education elementary school in Taiwan, Dora was diagnosed with muscular dystrophy and had two years computer experience for the purposes of education and leisure activities. The fact that she could only use her left index finger to control conventional keyboards and that she was not familiar with the QWERTY layout made it a difficult task for her to produce effective interaction with computers. Nonetheless, with instructions, she could type English letters; however, it remained challenging for her to perform the *Shift* key to type capital letters and punctuation marks.

3.2.2 Intervention

In the process of developing appropriate computer access for the participant, her characteristics should be taken into consideration. Prior to the intervention, the motor controls of Dora were assessed by the scale of computer access assessment [6]. The authors suggested Dora use a numeric keypad as an adaptive device (as shown in Figure 4).

The alternating treatments design, a type of single-subject design [7], was employed to compare the effectiveness of Dora's original text entry system with ETE system. Accuracy and speed performance were explored in the performance of text entry task.

Numbers of correct characters per minute (CPM) in each session represented the speed. Accuracy referred to the percentage of correct keystrokes in total trials. The intervention included three phases: 1) baseline, 2) comparison, and 3) final.

In the comparison phase, the participant was required to administer three sessions per week, each session consisting 45 minutes. To reduce the sequence effect of the two systems, the sequence of the devices alternated in each session. The 45-minute session started with a 5-minute practice for the first text entry system, and then she took a test in the following 5 minutes. Afterwards, she took a 5-minute break. Next, the participant went to the next text entry system, repeating the same process as was in the first method. Finally, in a sequential manner, the participant first conducted a 5-minute test using the first device and then the second device, with a five-minute interval break in between. The comparison phase terminated when the performance of both systems ceased to improve over 15% in three continuous sessions.

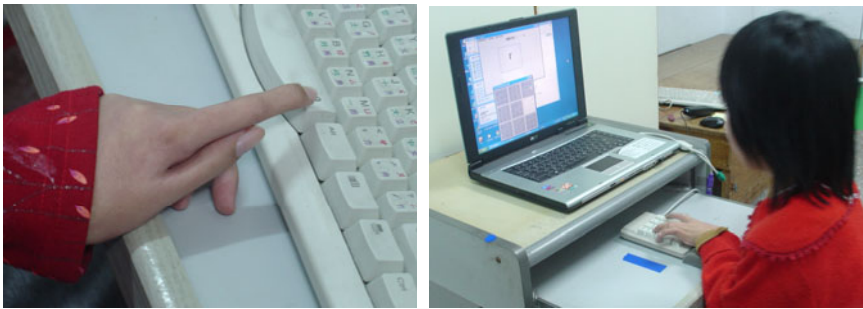


Fig. 4. The settings of the two text entry systems for Dora

The experiment used an Acer laptop with a Pentium M 730 Processor 1.5G, 512 MB of RAM, and a 15-inch monitor with 1024x768 pixels for the two keyboarding conditions. The text entry system and a keyboarding performance evaluation (KPE) program were installed in the laptop. The KPE program was utilized to generate the target stimuli and to collect the keystroke and time data. It displayed one English letter or punctuation mark on the screen at one time. Dora was instructed to press the matching key on the QWERTY keyboard or ETE system as quickly as possible. When the character was generated by pressing the key on either keyboard, the next target character displayed. The KPE system recorded the response time and accepted the selection without feedback. The KPE system automatically ended when the time was up.

3.2.3 Results

The CPM performance for Dora was shown in Figure 5. The first trial of the QWERTY keyboard was 17.3 CPM in the baseline. However, the speed increased to 26.9 CPM by the 8th trial. The CPM mean in the baseline was 24.5, showing a gradual improvement in performance ($Z=2.52$, $p<.01$). However, the last three points showed practically zero slope, which indexed that Dora entered the comparison phase.

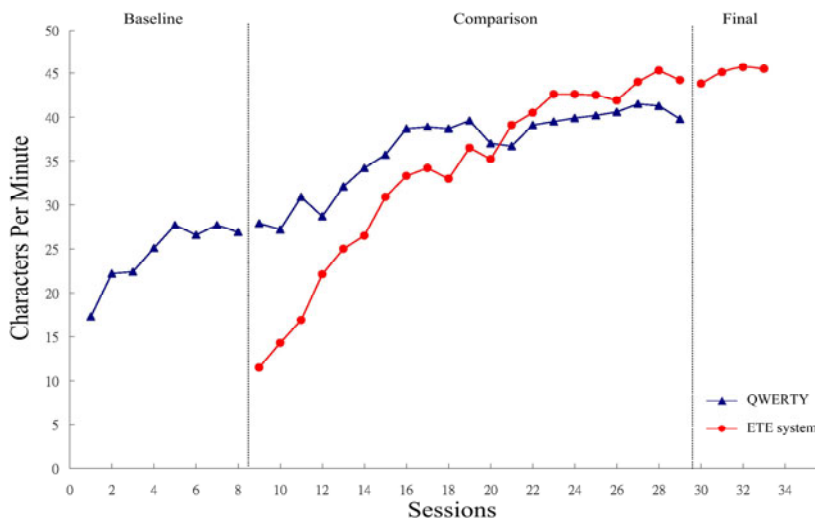


Fig. 5. The typing speed of the two text entry systems during the three phases for Dora

In the comparison phase, the CPM continued to improve from 27.9 to 41.5 CPM with positive trend. The mean of this phase was 36.6 CPM. The slope changed significantly ($z=4.45$, $p<.01$). The mean for the last three trails (40.9 CPM) was 1.43 times as fast as the first three trials (28.7 CPM).

Dora generated 11.5 CPM during the first trial of the coded text entry system. Her performance displayed steady improvement and, by the 13th trial, she achieved the rate of 39.1 CPM on D3 on-screen keyboard, reflecting that her performance enhanced by D3 keyboard exceeded that by the QWERTY keyboard. The mean of this phase was 33.4 CPM. The mean of the last three trials (44.5 CPM) was 3.13 times as fast as the first three trials (14.2 CPM). The slope of the phase changed significantly ($z=4.68$, $p<.01$) for D3 on-screen keyboard. Since the final three trials were stable, the comparison phase was terminated at session 29. Although the mean of the first three trials of coded text entry system was only 49.5% of the QWERTY, the mean of the last three trials was 1.09 times as fast as that of the QWERTY. In the final phase, Amy typed 45.0 CPM while the performance was stable and no trend was observed.

Dora's accuracy in using both keyboards remained consistent across the phases. As Figure 6 indicates, the mean for accuracy was 97.7% during baseline. The trend and slope of the phases were virtually zero. The accuracy for QWERTY keyboarding during the comparison phase remained high (mean=98.3%). The mean of the first three trials was 97.6% and the last three trails was 98.0%. The change of the slope was insignificant (Z value of C -statistic=.84).

For D3 on-screen keyboard, the accuracy of the first trial was 95.9% during the comparison phase. The mean of the first three trails was 97.6% while the last three trials was 98.7%. The mean of the phase was 98.3% with no significant slope change ($Z=-.35$). The accuracy of D3 on-screen keyboard maintained high in the final phase (mean=98.1%).

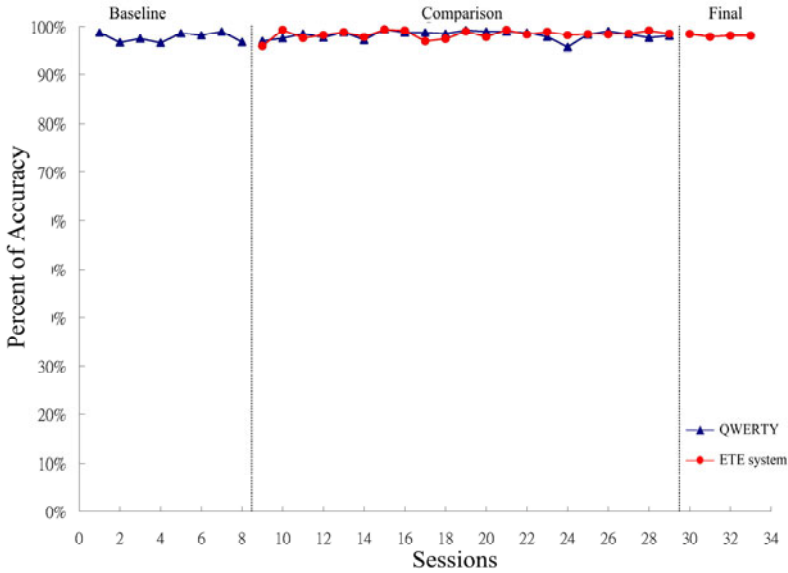


Fig. 6. The accuracy of the two text entry systems during the three phases for Dora

4 Discussions and Conclusions

The current study introduced an innovative text entry system that not only integrated both an on-screen keyboard and an encoding system, but also embedded a layout adaptation evaluation program which granted the options of the on-screen keyboard setting and a visual prompt strategy to encoding text entry system. The effects of the application on two clients of varied disabilities demonstrated the benefits of the embedding design.

Based on the findings of this study, three conclusions can be drawn: 1) Embedding visual prompts could facilitate the user experience as he/she employed encoding text entry system to successfully learn mapping codes of each character. 2) The layout adaptation evaluation program could assist the clinical professionals in detecting the proper size of the on-screen keyboard for those who could select the key directly. 3) The single integrated system could meet the diverse needs of the clients with severe physical disabilities.

Design for Dynamic Diversity is an ideal philosophy for device design. In compliance with the philosophy, it is essential that the user's performance be manually evaluated and the setting be adjusted, which creates complex tasks for clinic professionals. Based on the findings of this research, it is advised that the text entry system embed the auto detection function which could record the user's performance and then accordingly provide adjusting suggestions. Besides, although the visual prompts are effective for use of ETE system without requiring users to learn mapping codes in advance, the effect of memorizing codes was not explored in this paper. The future system might provide fading strategy after the user could type correctly with visual prompts further.

References

1. Cook, A.M., Polgar, J.M.: *Cook and Hussey's Assistive technologies: Principles and practice*, 3rd edn. Mosby, Baltimore (2008)
2. August, S., Weiss, P.L.: A human factors approach to adapted access device prescription and customization. *Journal of Rehabilitation Research Development* 29(4), 64–77 (1992)
3. Gregor, P., Newell, A.F., Zajicek, M.: Designing for dynamic diversity: making accessible interfaces for older people. In: *Proceedings of the 2001 EC/NSF workshop on Universal accessibility of ubiquitous computing*, pp. 151–156 (2001)
4. Fitts, P.M., Peterson, J.R.: Information capacity of discrete motor responses. *Journal of Experimental Psychology* 67(2), 103–112 (1964)
5. Lin, Y.L., Yeh, Y.M., Chen, M.C., Wu, Y.P., Wang, H.: A flexible on-screen keyboard: dynamically adaptive for individuals' needs. In: Stephanidis, C. (ed.) *HCI 2007. LNCS*, vol. 4556, pp. 371–379. Springer, Heidelberg (2007)
6. Wu, T.F., Meng, L.F., Wang, H.P., Wu, W.T., Li, T.Y.: Computer access assessment for persons with physical disability: A guide to assistive technology interventions. In: Miesenberger, K., Klaus, J., Zagler, W.L. (eds.) *ICCHP 2002. LNCS*, vol. 2398, pp. 204–211. Springer, Heidelberg (2002)
7. Alberto, P.A., Troutman, A.C.: *Applied behavior analysis for teachers*, 5th edn. Macmillan Publishing Co., New York (1999)

Qanti: A Software Tool for Quick Ambiguous Non-standard Text Input

Torsten Felzer¹, Ian Scott MacKenzie²,
Philipp Beckerle¹, and Stephan Rinderknecht¹

¹ Institute for Mechatronic Systems, Technische Universität Darmstadt,
Darmstadt, Germany

² Department of Computer Science and Engineering, York University,
Toronto, Canada M3J 1P3

{felzer, beckerle, rinderknecht}@ims.tu-darmstadt.de,
mack@cse.yorku.ca

Abstract. This paper introduces a single-key text entry application for users with severe physical impairments. The tool combines the idea of a *scanning ambiguous keyboard* (which promises unusually high entry rates) with *intentional muscle contractions* as input signals (which require much less physical effort compared to key presses). In addition to the program architecture, the paper presents the results of several evaluations with participants with and without disabilities. An entry speed of 6.59 wpm was achieved.

Keywords: Human-computer interaction, scanning, ambiguous keyboards, intentional muscle contractions.

1 Introduction

Persons with physical disabilities are often unable to use a standard keyboard when interacting with a computer. Therefore, text-based communication typically involves an alternative input method, often relying on *scanning*. With scanning, the computer “suggests” an action (e.g., the selection of a character) by cyclically highlighting through a set of options with a *scan delay* τ (typically in the range 0.5 s to 2.0 s). The user selects the highlighted option by issuing a *selection marker* (which means, for example, pressing a single key or actuating a physical switch).

Unfortunately, while demanding less physical effort, scanning systems are usually prohibitively slow – considerably slower than keyboard-based text entry. Therefore, entering text is often cumbersome for physically disabled users with the consequence that communication methods requiring fast typing (e.g., Internet chat) are not possible. A fast one-key input device could tremendously improve the quality of life for physically disabled users by empowering them to participate in a world of communication opportunities formerly denied to them.

This work builds the idea of a *scanning ambiguous keyboard* (SAK) – as presented in ref. [1] – into a fully-implemented system. The SAK approach considers

multiple selections in a single scan step, which promises unusually high entry rates. However, SAK designs generally work only for words in a known list – a *dictionary*. Our system addresses this using a spell mode to add new words to the dictionary. In addition, the scanning approach is combined with the input signals in [2], i.e., *intentional muscle contractions*, to further reduce the required physical effort.

The resulting software application – dubbed *Qanti* for “Quick Ambiguous Non-standard Text Input” – is a tool specifically for persons with severe physical disabilities, allowing the user to effortlessly enter text in an acceptable amount of time. After reviewing approaches to one-key text entry in the literature in section 2, the *Qanti* solution is described in section 3. The system’s practical effectiveness will be demonstrated in several small-scale usability studies in section 4. The paper concludes with a summary and a look at future work.

2 Known Approaches

Making computers accessible to physically disabled users is challenging for at least two reasons. First, using a computer is virtually impossible without entering text – and alternative text entry applications (with a small number of keys or input signals) are therefore needed (e.g., [3]). Second, reducing the number of keys to represent the 26 letters of the alphabet (plus certain “special” characters like space, dash, apostrophe or various punctuation) brings about special challenges [4,5].

To confront these challenges, numerous acceleration techniques supplementing the scanning idea [6] have been developed. A straightforward idea is to reduce the scan delay, as a shorter delay increases the speed. But there is a trade-off: very short delays cause erroneous or missed selections, *reducing* the entry rate. The approach of Simpson and Koester [7] tries to cope with this problem by dynamically adapting the scan delay (“on the fly”) based on user performance.

Another idea is to use *word completion* (e.g., [8]), where the computer tries to “guess” a word based on the characters entered so far. Displaying a list of candidate words and offering a mechanism to select the desired candidate early helps save “keystrokes” (i.e., selections). However, constantly having to inspect the candidate list also increases the *cognitive load* on the user.

As text entry progresses, selectable options might increase in likelihood compared to alternatives (according to syntactic or semantic considerations). The system described by Wandmacher et al. [9] takes advantage of this by dynamically reordering the scanned options, so that more likely options are selected faster.

The solution presented here reduces the number of steps in a scan cycle by associating more than one character to a single virtual key (also [10]). The actual words belonging to the key sequences are computed using dictionary-based disambiguation.

3 Software Tool

The text entry application *Qanti* is depicted in fig. 1. It divides the screen into four areas: the sequence selection area on the upper left, an output area on the

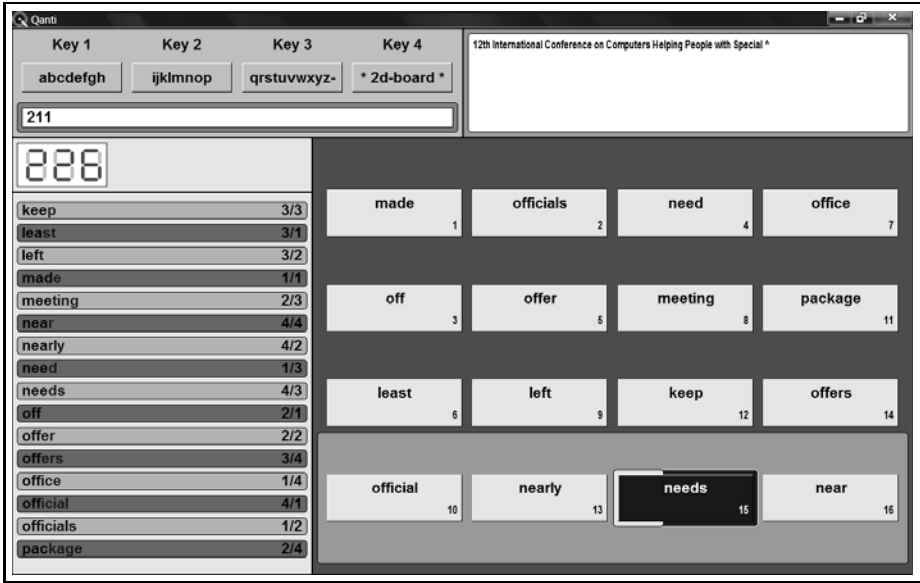


Fig. 1. Specialized writing tool *Qanti*

upper right showing the text entered so far, an information area on the lower left, and a large 4×4 scanning board on the lower right.

3.1 Typical Operation

Entering a word typically involves three phases: a sequence selection phase, a candidate selection phase, and a finalization phase. First, the user enters a sequence consisting of the linearly scanned virtual keys in the sequence selection area by issuing intentional contractions as *selection markers*. Each of the first three keys is associated with approximately one third of the alphabet (in analogy to a phone keypad where most keys are associated with three letters). An important property of this phase is that the scanning timer is reset after each selection with the highlight persisting on the current key. As a consequence, the user can make multiple selections during a single scan step – which makes the method enormously fast.

The program responds to each selection of a “character key” by determining all candidate words matching (and completing!) the current sequence in a frequency-ordered dictionary. The 16 most frequent candidates are displayed in the information area (in alphabetic order) *and* on the virtual keys of the two-dimensional board (ordered by frequency). If the desired word is among those presented, the user can switch to the next phase by selecting the fourth virtual key in the sequence selection area. After the selection of the intended candidate (using conventional row-column scanning), the user is prompted to finalize the

word by choosing a *modifier* (which adds a space, comma, or period at the end or capitalizes the first character).

Note that the candidates are *not* placed in a “row-oriented” but in a “diagonal-oriented” order on the 2d-keyboard (seen from the button indexes in fig. 1). The reason for this is the selection time required using row-column scanning – for instance, selecting the third button in row 1 (which would be the position of candidate 3 in “row-oriented” order) takes *longer* than selecting the first button in row 2 (the “row-oriented” position of candidate 5)!

Finally, a feature of the software is its multilingual support. The user may choose either English or German. This choice affects the output strings used throughout the program, the pre-loaded dictionary, and the selectable characters, for instance, “umlauts” when using German.

3.2 Menu Mode

In addition to typical word selection, the user can enter a menu mode by selecting the fourth “sequence key” *without* prior selection of a “character key” (i.e., without any other selections in the same scan cycle). The menu options include resetting the current sequence, configuring the scan delay, or invoking a conventional row-column scanning keyboard for entering “out of dictionary” words.

Another option launches an automated *test protocol* which randomly selects phrases from a phrase set supplied to the program, displays the phrases in the output area, and records the time for the user to copy the phrases.

After starting the test, an additional timer is displayed (not shown in fig. 1) together with the phrase to be copied. The user then has the opportunity to study the phrase and “plan ahead”, because the timer is halted until the next input signal. When the user is finished transcribing the phrase, he/she can add a newline character. *Qanti* responds to that by computing the current entry speed – in *words per minute* (wpm) – etc., writing the statistics to disk, and displaying the next phrase to be copied (with the timer again suspended until the user issues another input signal).

3.3 Input Signals

As already mentioned, the software supports intentional muscle contractions as input signals. The operation of the program relies on *scanning* through a number of options, and to make a selection in this context, the user just issues tiny contractions of a single muscle of choice.

To detect those contractions, the activity of the muscle in question – indirectly recorded with the help of a piezo element – is constantly compared to a user-dependent threshold. Exceeding this *sensitivity* threshold triggers a contraction event and, thus, a selection.

The prototypical configuration – utilized in the evaluation studies in the next section – is tailored to the brow muscle, i.e., it involves a piezo pressed against the eyebrow by means of an elastic sports headband. The study participants

therefore merely needed to raise the eyebrow (or to *frown*), in order to produce selection markers.

4 Empirical Data

The program is particularly powerful concerning long and frequently used words. For instance, *Qanti* allows the user to enter the 14-character string “International ” (including the trailing space) in less than 12 seconds. This is of course a non-typical, very advantageous example, yet it shows that peak velocities of more than 1 character per second (i.e., more than 12 wpm) are possible in theory.

The practical aspects of *Qanti* were tested through three studies: a pilot study with one subject diagnosed with *Friedreich’s Ataxia* (FA), an attempted (but not completed) test with two subjects with *Spinal Muscular Atrophy* (SMA), and a usability study involving five able-bodied users. Each participant was male and between 22 and 39 years old.

The task of all eight subjects was to copy phrases – as quickly and accurately as possible (correcting errors if noticed) – presented at random from a phrase set supplied to the program. It was ensured that each unique word in the phrase set was also in the program’s dictionary. The phrase set consisted of 56 common German sayings and proverbs with an average length of 56.2 characters (ranging from 31 to 81) – some typical examples are shown in fig. 2.

Es wird nichts so heiß gegessen, wie es gekocht wird.
 Wächst das Gras im Januar, so wächst es schlecht im ganzen Jahr.
 Was du heute kannst besorgen, das verschiebe nicht auf morgen.
 Wer anderen eine Grube gräbt, fällt selbst hinein.
 Wovon man nicht sprechen kann, darüber muss man schweigen.

Fig. 2. Sample phrases used in the evaluations

The FA patient in the pilot study has used a wheelchair for more than 20 years and has considerable motor problems. However, his usual method of entering text involves a standard keyboard at a typing rate of typically 3 – 6 wpm. He has participated in a number of empirical studies concerning alternative text entry and intentional muscle contractions. He was even familiar with *Qanti*, so he was a perfect choice to do some initial tests and verify that the software runs smoothly (the purpose of the pilot study).

The task was to enter five blocks consisting of three phrases each, using the participant’s preferred scan delay (1000 ms). To avoid over-exertion, rest intervals of at least five minutes were inserted between blocks, and the test protocol described in the previous section provided for additional pauses between individual phrases. The average entry speed achieved in the five blocks is illustrated in fig. 3a).

The participant indicated that he was quite satisfied with this result (around 2.5 wpm), which was close to the lower bound of his usual typing speed – yet

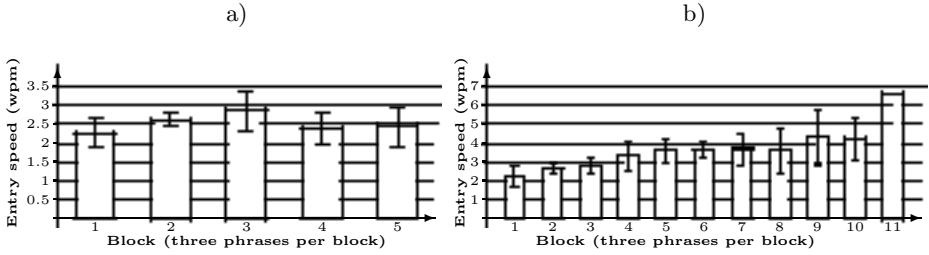


Fig. 3. Entry speed (in wpm) achieved in the evaluation studies: *a)* Pilot study with one participant with FA (using a scan delay of 1000 ms); *b)* usability study with five able-bodied participants (the scan delay was 1000 ms in block 1 and decreased by 50 ms per block, finishing at 500 ms in block 11)

requiring only a fraction of the physical effort (raising the eyebrow versus manual typing). In addition, he proceeded to achieve even higher rates with extended training.

After the successful pilot study, the program was considered ready for evaluation by members of the target population. Two participants with SMA (both not able to enter text manually at all) volunteered for this evaluation. The plan was to explain the program and to have them try the input concept in two initial sessions (one for each) and to have them copy an indetermined number of phrases (as many as they would like) in timed test sessions.

The participants were both able to produce intentional muscle contractions and to use those to operate the program in the individual initial session. Each of them proved this by spelling his first name with the row-column scanning keyboard as well as adding it to the program's dictionary and subsequently retrieving it using the three-phase word selection process detailed above. Unfortunately, due to medical problems which emerged on that day (not caused by the use of *Qanti*), the actual tests had to be canceled.

The participants for the third study were recruited among the students of the Technische Universität Darmstadt. Again, the testing was divided into two parts for each subject. At first, the participants were introduced to the software. Afterwards, they engaged in a session containing up to 11 blocks. Each block consisted of three sentences taken from the phrase set mentioned above. The difference between the successive blocks was the scan delay decreasing by 50 ms per block, starting at 1000 ms. The users' task was to copy the presented sentences and proceed from one block to the next until they were not able to keep up with the scan delay.

Fig. 3b) shows the entry speeds of the participants plotted against the blocks, which corresponds to a decreasing scan delay. The differences between the single sentences within the blocks of the individual participants are not shown – the values in the chart represent an overall average. It is evident in the chart that the entry speed is increasing with decreasing scan delay. Starting at 2.22 wpm

at 1000 ms, the average speed increases to 6.59 wpm at 500 ms (block 11) due to the shorter waiting intervals. At the same time, the standard deviation (represented by the error bars in the figure) is rising with the blocks. This is caused by the increased concentration required. Some participants had problems in concentrating at lower scan delays, while others were able to raise their entry speed even more, because they were able to retain their concentration. In addition to that, the participants aborted the test when they were not able to keep up with the delay. The first participant aborted the test after accomplishing the 700 ms block, the next two after the 650 ms block. The remaining two aborted after 550 ms and 500 ms. So the last block of the chart does not have an error bar, because it relies on the results of one participant only.

Another interesting result is the number of sentences entered correctly. This number increased with the decreasing scan delay at the beginning of the study due to learning effects of the participants (increasing from 2.4 correct sentences at 1000 ms to 3.0 correct sentences at 850 ms). It decreased to 2.0 correct sentences at 650 ms, because the participants had to concentrate more to keep up with the scan delay. At delays between 600 ms and 500 ms, it was increasing again due to the fact that only the most proficient participants were able to continue with short delays.

5 Conclusion and Future Work

A tool for fast but effortless text entry designed for persons with severe physical disabilities has been introduced. The software makes use of *scanning* and thus allows users to enter text with a single input signal – an *intentional muscle contraction*. Entering a word typically involves producing a sequence of three “character keys” – each ambiguously representing about one-third of the alphabet – and selecting the desired candidate from a list of words matching or extending the entered sequence within the program’s *dictionary*.

Empirical data collected in several evaluation studies confirm the practical usefulness of the approach. The resulting entry rates ranged from 2.5 to 6.5 wpm, which is quite competitive for a scanning program, given that *Qanti* is a ready-to-use system supporting error correction and out-of-dictionary words.

The achievable entry rate heavily depends on the dictionary used by the system, so updating this dictionary yields a huge potential for future work. In this respect, it is planned to automatically extract a personalized corpus from text files already written by the user. Another idea for future work is to integrate *Qanti* into higher-level software – for example, the scanning-based telephony application of ref. [11] – as a universal text entry sub-system.

Acknowledgments. This work is partially supported by DFG grant FE 936/3-2 “The AID package – An Alternative Input Device based on intentional muscle contractions”.

References

1. MacKenzie, I.S., Felzer, T.: SAK: Scanning ambiguous keyboard for efficient one-key text entry. *ACM Transactions on Computer-Human Interaction (TOCHI)* (in Press)
2. Felzer, T., Nordmann, R.: Evaluating the hands-free mouse control system: An initial case study. In: Miesenberger, K., Klaus, J., Zagler, W.L., Karshmer, A.I. (eds.) *ICCHP 2008*. LNCS, vol. 5105, pp. 1188–1195. Springer, Heidelberg (2008)
3. Baljko, M., Tam, A.: Motor input assistance: Indirect text entry using one or two keys. In: *Proc. ASSETS 2006*, pp. 18–25. ACM Press, New York (2006)
4. Cardinal, J., Langerman, S.: Designing small keyboards is hard. *Theor. Comput. Sci.* 332(1-3), 405–415 (2005)
5. MacKenzie, I.S., Tanaka-Ishii, K.: Text entry with a small number of buttons. In: MacKenzie, I.S., Tanaka-Ishii, K. (eds.) *Text entry systems: Mobility, accessibility, universality*, pp. 105–121. Morgan Kaufmann, San Francisco (2007)
6. Felzer, T., Strah, B., Nordmann, R.: Automatic and self-paced scanning for alternative text entry. In: *Proc. IASTED Telehealth/AT 2008*, pp. 1–6. IASTED (2008)
7. Simpson, R.C., Koester, H.H.: Adaptive one-switch row-column scanning. *IEEE Transactions on Rehabilitation Engineering* 7(4), 464–473 (1999)
8. Felzer, T., Nordmann, R.: Speeding up hands-free text entry. In: *Proc. CWUAAT 2006*, pp. 27–36. Cambridge University Press, Cambridge (2006)
9. Wandmacher, T., Antoine, J.Y., Poirier, F., Départe, J.P.: Sibylle, an assistive communication system adapting to the context and its user. *ACM Trans. Access. Comput.* 1(1), 1–30 (2008)
10. Belatar, M., Poirier, F.: Text entry for mobile devices and users with severe motor impairments: handiglyph, a primitive shapes based onscreen keyboard. In: *Proc. ASSETS 2008*, pp. 209–216. ACM Press, New York (2008)
11. Felzer, T., Beckerle, P., Rinderknecht, S.: How to make a telephone call when you cannot operate a telephone. In: Langdon, P., Clarkson, J., Robinson, P. (eds.) *DESIGNING INCLUSIVE INTERACTIONS: Inclusive Interactions Between People and Products In Their Contexts Of Use*, pp. 177–186. Springer, Heidelberg (2010)

A Prototype Scanning System with an Ambiguous Keyboard and a Predictive Disambiguation Algorithm

Julio Miró-Borrás, Pablo Bernabeu-Soler, Raul Llinares, and Jorge Igual

Departamento de Comunicaciones, Universidad Politécnica de Valencia
Plaza Ferrandiz-Carbonell s/n, 03801 Alcoy, Spain
{jmirobo,pbernabe,rllinares,jigual}@dcom.upv.es

Abstract. This paper presents a prototype single-switch scanning system with a four-key ambiguous keyboard. The keyboard consists of only three keys for letters and an additional key for editing and extra functions. The letter-to-key assignment is based on the shapes of lowercase consonants, and vowels are alphabetically distributed among the keys. An adaptation of the "one-key with disambiguation" algorithm has been used to increase text entry speed. The system has been implemented for Spanish language.

Keywords: Ambiguous Keyboards, Scanning Systems, Text Entry, AAC.

1 Introduction

People with severe motor disabilities usually require scanning systems (ScS) to enter text into a computer or even to communicate with other people. Traditional ScS are based on a scan matrix with one character per cell. We present in this paper a single-switch ScS prototype with an ambiguous virtual keyboard and a predictive disambiguation algorithm. The keyboard layout is based on what we have called CGA (Character Graphic Association), a new, easy-to-learn way to assign letters to keys depending on the consonants' shape. The keyboard presents only 3 keys for letters and one extra key for editing and extra functions. Predictive disambiguation algorithms are typically used in mobile phones and we adapted the "one-key with disambiguation" algorithm to a single switch ScS with ambiguous keyboards, as described in [1]. Using this algorithm requires one keystroke per letter and another to select the desired word in the list of proposed words.

The use of ambiguous keyboards in ScS is not new. In the literature we can find several ScS using different layouts styles: abc layouts such as SAK [2], optimized layouts such as AKKO [3] or even graphical layouts such as UniGlyph [4]. Other common ambiguous keyboards used in cell phones follow QWERTY ordering. The user interface and the use of extra keys are different for each proposal.

Having several letters on each key makes it necessary to use a method to indicate to the system the intended character or word the user wants to enter. The most used disambiguating proposals in mobile phones are multitap, two-keys, and one-key with disambiguation, as described in [5]. The first two are non-predictive methods and the last one is a predictive method based on a dictionary. Most ScS with ambiguous

keyboards use predictive methods. In [6] we also propose an adaptation of the multi-tap method to ScS.

2 The Prototype

2.1 CGA Layouts

CGA represents a new way to assign letters to keys and allows us to generate keyboard families with two, three and four keys. We group the consonants that share the same position with respect to the lines on ruled paper, when letters are handwritten on the lines. As shown in Fig. 1, these groups are: (1) consonants exceeding the top line (b, d, f, h, k, l, t); (2) consonants between the lines (c, m, n, r, s, v, w, x, z); and (3) consonants exceeding the bottom line (g, j, p, q, y). This classification of letters is quite intuitive and very easy to learn and remember. In order to improve performance, vowels are treated in a different way: (1) only consecutive alphabetically-ordered vowels can be assigned to a key; and (2) the keyboard must be able to display the five vowels in strict alphabetical order. The layout in the prototype, named CGA3-122¹, is shown in Fig. 2 and follows our CGA criteria.

2.2 Disambiguation Algorithm and User Interface

The disambiguation algorithm used is the Word Level Disambiguation Mode, similar to the "one-key with disambiguation" used in mobile phones but adapted to the single-switch ScS, as described in [1]. It uses neither a separate key for space nor one for the "next" function typically used in the disambiguation process. Both keys are replaced by a combination of automatic and inverse scanning and the allowance of long and short presses of the switch. Text entry is accomplished through two phases. The first phase (or cell selection phase) is equivalent to the action of pressing the keys on a "physical" keyboard, but using automatic scanning. The cursor advances to the next cell each scan period using automatic scanning and the user presses the switch once for each letter of a word, except the last one. Fig. 2 shows the cell selection phase window, and the user is typing the Spanish word "pescado" (fish). The sequence of cells selected (code sequence) is 122213² for the first 6 letters of that word. Entering the last letter is done by keeping the switch pressed when the desired cell is highlighted. This tells the system that this was the last letter in the word, and a space will be added after the word disambiguation. In the later example, the last letter for "pescado" is "o", so the user keeps the switch pressed when the cell on the right is highlighted (corresponding to code 3). Then the second phase (disambiguation phase) starts (Fig. 3). The words in the dictionary that share the same sequence of selected cells (code) are shown one after another, the most probable one first, by using inverse scanning on a wide key replacing the CGA layout (Fig. 4). The selection of a word is done by releasing the switch when the desired word appears on the key. Then the

¹ CGA3-122 is the given name for the 122nd layout in our three-key family list of keyboards.

² Numbers represent the selected cells for each letter of the word: 1 for the cell on the left, 2 for the middle one, and 3 for the cell on the right.

word is accepted and a new word can be typed again. The word dictionary consists of the 11,000 most frequently used words in Spanish.

The prototype implements basic editing functions such as "delete character" or "delete word". In the cell selection phase window when the key labeled as "Borrar" ("delete" in Spanish) is selected, another window presents functions to delete a character or a word, and also implements a text to speech characteristic to read aloud the typed phrase.



Fig. 1. Icons representing groups of consonants

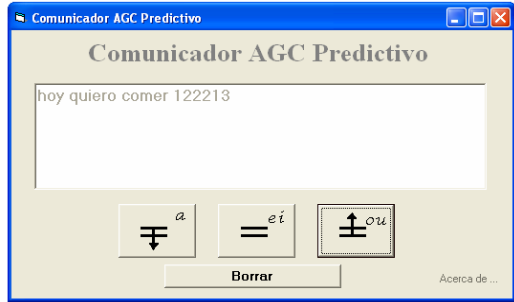


Fig. 2. Main window (cell selection phase)

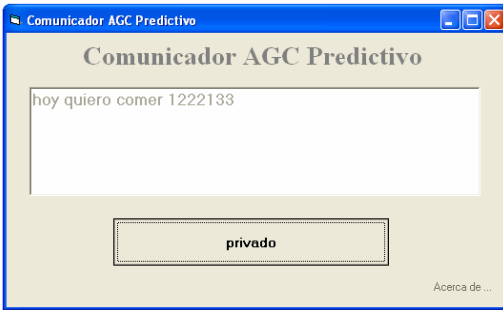


Fig. 3. Disambiguation phase

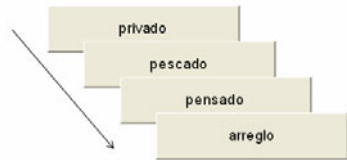


Fig. 4. Word Selection

3 Discussion

The use of CGA layouts is very convenient for people with motor disabilities and presents the following advantages: it is easy to learn and remember; simple icons represent the set of consonants in each cell; and keyboards have very few keys. One of the advantages of CGA keyboards is the ability to generate different layouts simply by assigning vowels to other keys. This way it is possible to find a layout with better performance when using another linguistic model, another disambiguation method, or even for other languages with Latin alphabets. Reduced devices implementing predictive algorithms usually require the use of a key for the "space" character and another

to select the desired word ("Next" key). The proposed algorithm avoids the use of both keys although other implementations with other disambiguation algorithms may require them. Although this prototype has been implemented for the Spanish language, performance prediction for expert users has been done with the linguistic model and parameters described in [1], concluding a text entry speed of 16.6 wpm [7].

4 Conclusions and Planned Lines of Research

This prototype ScS combines an easy-to-learn ambiguous keyboard and a predictive disambiguation method. The expected advantages are a low learning time and a faster text entry rate mainly due to the reduction of the number of scan cycles to reach the cells during scanning and the use of prediction.

We are currently working in two areas: (a) adapting a character level algorithm similar to multitap which is described in [1] and [6]; and (b) incorporating other characters such as numbers and punctuation items. We have also planned to conduct experimental tests with users.

Acknowledgments. This study has been partly funded by Cátedra Telefónica - UPV.

References

1. Miró-Borrás, J., Bernabeu-Soler, P.: Text Entry in the E-Commerce Age: Two Proposals for the Severely Handicapped. *J. Theor. Appl. Electron. Commer. Res.* 4, 101–112 (2009)
2. MacKenzie, I.S.: The one-key challenge: Searching for a fast one-key text entry method. In: *ACM Conference on Computers and Accessibility*, pp. 91–98 (2009)
3. Harbusch, K., Kühn, M.: Towards an adaptive communication aid with text input from ambiguous keyboards. In: *10th Conference on European Chapter of the Association for Computational Linguistics*, pp. 207–210 (2003)
4. Belatar, M., Poirier, F.: Text Entry on Mobile Devices for Users with Severe Motor Impairments: HandiGlyph, a Primitive Shapes Based Onscreen Keyboard. In: *10th Annual ACM SIGACCESS Conference on Assistive Technologies*, pp. 209–216 (2008)
5. MacKenzie, I.S., Tanaka-Ishii, K.: Text entry using a small number of buttons. In: *Text Entry Systems: Mobility, Accessibility, Universality*, pp. 105–121. Morgan Kaufmann, San Francisco (2007)
6. Miró, J., Bernabeu, P.A.: Text entry system based on a minimal scan matrix for severely physically handicapped people. In: Miesenberger, K., Klaus, J., Zagler, W.L., Karshmer, A.I. (eds.) *ICCHP 2008. LNCS*, vol. 5105, pp. 1216–1219. Springer, Heidelberg (2008)
7. Miró-Borrás, J., Bernabeu-Soler, P., Llinares, R., Igual, J.: An Ambiguous Keyboard based on "Character Graphical Association" for the Severely Physically Handicapped. In: *ICCHP 2010*, Springer, Heidelberg (2010)

An Ambiguous Keyboard Based on "Character Graphical Association" for the Severely Physically Handicapped

Julio Miró-Borrás, Pablo Bernabeu-Soler, Raul Llinares, and Jorge Igual

Departamento de Comunicaciones, Universidad Politécnica de Valencia
Plaza Ferrandiz-Carbonell s/n, 03801 Alcoy, Spain
{jmirobo,pbernabe,rllinares,jigual}@dcom.upv.es

Abstract. Ambiguous keyboards can be used instead of scan matrices in Scanning Systems in order to increase the text entry rate. We present a novel alternative to assigning letters to keys by taking into consideration the shapes of lowercase letters in order to create simple layouts, leading to families of keyboards with 2, 3 and 4 keys. We have chosen the best 3-key layout using a single switch scanning system and a word level disambiguation algorithm with a 10,911 word dictionary and we have tested it. We predict 16.6 wpm for an expert user with a 0.5 second scan period.

Keywords: Ambiguous Keyboards, Scanning Systems, Text Entry, AAC.

1 Introduction

Keyboards with fewer keys than characters (ambiguous keyboards) can be used instead of scan matrices in Scanning Systems (ScS) in order to increase the text entry rate. Different arrangements of characters produce different layouts, trying to combine ease of learning and performance. We present a novel alternative to assigning letters to keys by taking into consideration their graphical characteristics in order to create simple layouts. We named it CGA (Character Graphical Association). We then analyze the performance of a 3-key CGA layout named CGA3-122¹ when used in a single-switch ScS with predictive disambiguation as described in [1].

The use of ambiguous keyboards in ScS is not new. In the literature we can find several ScS using different layouts styles: abc layouts such as SAK [2] (Fig. 1), optimized layouts such as AKKO [3] (Fig. 2) or even graphical layouts such as UniGlyph [4] (Fig. 3). Other common ambiguous keyboards used in cell phones follow QWERTY ordering. The user interface and the use of extra keys are different for each proposal. Having several letters on each key makes it necessary to use a procedure to indicate to the system the intended character or word the user wants to enter. The most used disambiguating proposals in mobile phones are multitap, two-keys, and one-key with disambiguation, as described in [5]. The first two are non-predictive methods and the last one is a predictive method based on a dictionary. Most ScS with

¹ CGA3-122 is the given name for the 122nd layout in our 3-key family list of keyboards.

ambiguous keyboards use predictive methods. In [6] we also propose an adaptation of the multitap method to ScS.

1.1 CGA Layouts

CGA represents a new way to assign letters to keys and allows us to generate keyboard families with two, three and four keys. We group the consonants that share the same position with respect to the lines on ruled paper, when letters are handwritten on the lines (Fig. 4). These groups are: (1) consonants exceeding the top line (b, d, h, k, l, t); (2) consonants between the lines (c, m, n, r, s, v, w, x, z); and (3) consonants exceeding the bottom line (g, j, p, q, y). This classification of letters is quite intuitive because it is easy to associate a letter with its drawing when using appropriate icons on top of the keys (Fig. 5). When handwritten, the "f" letter exceeds both lines. Performance reasons should be taken into consideration for choosing group 1 or 3 (Fig. 5).



Fig. 1. SAK layout

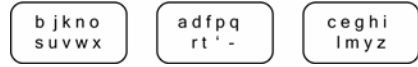


Fig. 2. AKKO layout



Fig. 3. UniGlyph layout

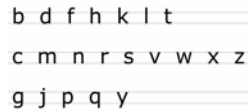


Fig. 4. Groups of consonants

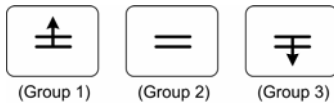


Fig. 5. Icons representing groups of consonants

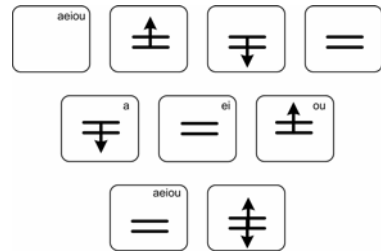


Fig. 6. Several CGA layouts

Vowels should be treated in a different way because all of them belong to group 2 and the sum of their frequencies in English represents more than 26% of all possible character frequencies. In order to improve performance, it is necessary to distribute the vowels among different keys. The following requirements should be fulfilled in order to make it easier for users to learn the layout: (1) only consecutive alphabetically-ordered vowels can be assigned to a key; and (2) the keyboard must be able to display the five vowels in strict alphabetical order. Fig. 6 shows some examples of four, three and two-key layouts with the vowels printed on the keys. CGA3-122 is the 3-key layout shown in Fig. 6, with "f" letter in group 1.

2 Methods

The research is done using the word-level disambiguation proposal in [1]. Using a combination of automatic/inverse scanning and short/long keypresses avoids the need for separate keys for the space character and the next function (typical in predictive disambiguation modes). The operation of this proposal is composed of two phases. The first one is the cell selection phase. Using automatic scanning, the user presses the switch when the desired cell is highlighted. This is done for all the characters in a word except for the last one. For the last letter, the user presses and keeps the switch pressed, indicating to the system that the word is finished. Then, the disambiguation phase starts, presenting the suggested words one after another using inverse scanning. When the desired word is shown, the user releases the switch and the word followed by a space is entered into the system. The word-frequency list used is the same as described in [1] and also the order in which cells are scanned according to their probabilities. Considering keystrokes both the action of pressing the switch and cursor movements, a parameter that comprises all kinds of keystrokes is n_W or average weighted number of keystrokes per word (Equation 1) [1], where $n_W(w)$ is the weighted number of keystrokes in word w , and $p(w)$ its probability. Additionally, w_S , w_C , and w_L are the weights for the scan cycles, the short and the long presses of the switch respectively, and n_S , n_C and n_L are the average values of the respective number of keystrokes.

$$n_W = \sum_w n_W(w) \cdot p(w) = w_S \cdot n_S + w_C \cdot n_C + w_L \cdot n_L. \quad (1)$$

Text entry rate in wpm (words per minute) can be obtained from Equation 2 [1], where T is the scan period.

$$wpm = 60 / (T \cdot n_W). \quad (2)$$

3 Results and Discussion

The estimated text entry rate for the proposed keyboard CGA3-122 is 16.6 wpm, taking into consideration the following assumptions: a scan period of 0.5 seconds and the weights used in [1], i.e. $w_S=1$, $w_C=0.5$ and $w_L=1$. If we consider the ScS with the same disambiguation algorithm but we replace CGA3-122 layout with other 3-key layouts and keep all the parameters constant, we obtain the following rates: (15.1 wpm for the SAK layout), (15.6 wpm for the AKKO layout) and (15.6 wpm for the UniGlyph layout). CGA3-122 presents the best performance among the 3-key layouts studied for this particular ScS. This is mainly due to the reduced number of keystrokes in phase 1 compared to other layouts that mainly try to minimize the number of keystrokes in the disambiguation phase [7].

The use of CGA layouts is very convenient for people with motor disabilities and presents the following advantages: it is easy to learn and remember; simple icons represent the set of consonants in each cell; and keyboards have very few keys. The ability to generate different layouts allows us to find a layout with better performance

when using another linguistic model, another disambiguation method, or even for other languages with Latin alphabets.

4 Lines of Research

We have planned several lines of research in two main areas. The first area is related to using other disambiguation modes, and the incorporation of other characters such as numbers and punctuation items. The second area is focused on conducting some experimental tests with users in order to study the system usability and to validate the analytical results.

Acknowledgments. This study has been partly funded by Cátedra Telefónica - UPV.

References

1. Miró-Borrás, J., Bernabeu-Soler, P.: Text Entry in the E-Commerce Age: Two Proposals for the Severely Handicapped. *J. Theor. Appl. Electron. Commer. Res.* 4, 101–112 (2009)
2. MacKenzie, I.S.: The one-key challenge: Searching for a fast one-key text entry method. In: *ACM Conference on Computers and Accessibility*, pp. 91–98 (2009)
3. Harbusch, K., Kühn, M.: Towards an adaptive communication aid with text input from ambiguous keyboards. In: *10th Conference on European Chapter of the Association for Computational Linguistics*, pp. 207–210 (2003)
4. Belatar, M., Poirier, F.: Text Entry on Mobile Devices for Users with Severe Motor Impairments: HandiGlyph, a Primitive Shapes Based Onscreen Keyboard. In: *10th Annual ACM SIGACCESS Conference on Assistive Technologies*, pp. 209–216 (2008)
5. MacKenzie, I.S., Tanaka-Ishii, K.: Text entry using a small number of buttons. In: *Text Entry Systems: Mobility, Accessibility, Universality*, pp. 105–121. Morgan Kaufmann, San Francisco (2007)
6. Miró, J., Bernabeu, P.A.: Text entry system based on a minimal scan matrix for severely physically handicapped people. In: *Miesenberger, K., Klaus, J., Zagler, W.L., Karshmer, A.I. (eds.) ICCHP 2008. LNCS*, vol. 5105, pp. 1216–1219. Springer, Heidelberg (2008)
7. Miró-Borrás, J., Bernabeu-Soler, P., Llinares, R., Igual, J.: Ambiguous Keyboards and Scanning: the Relevance of the Cell Selection Phase. In: *Gross, T., Gulliksen, J., Kotzé, P., Oestreicher, L., Palanque, P., Prates, R.O., Winckler, M. (eds.) INTERACT 2009, Part II LNCS*, vol. 5727, pp. 1–4. Springer, Heidelberg (2009)

The Deaf and Online Comprehension Texts, How Can Technology Help?

Simona Ottaviano, Gianluca Merlo, Antonella Chifari, Giuseppe Chiazese,
Luciano Seta, Mario Allegra, and Valentina Samperi

Italian National Research Council - Institute for Educational Technologies
Via Ugo La Malfa 153, 90146 Palermo, Italy
gym2learn@itd.cnr.it

Abstract. This paper focuses on metacognition and reading comprehension processes in hearing impaired students, as deafness causes a number of linguistic difficulties, affecting the comprehension and production of written text and reducing interest in reading. Then a web tool called Gym2Learn is described and the results of a single-subject-design are discussed. In conclusion, the effects of this system on deaf students' learning performance are highlighted.

Keywords: Hearing impaired, Web tool, Metacognition.

1 Introduction

Deaf people communicate using sign language which exploits a visual-gestual communication channel, particularly suited to their experiences of the world. Since the deaf have never been immersed in the so called "sound bath", the mixture of sounds, words and speech in the surrounding environment, they are lacking in knowledge about the "acoustic" world. They make up for this with sight, which also serves as a tool for acquiring knowledge, decoding reality and communicating. Consequently, the deaf mostly learn through vision and they also think using visual images and mental representations of their observations, opinions and views.

The use of spoken language does not reflect the style of knowledge or communication of deaf people, and this leads to severe difficulties in producing and understanding written language. Their inability to use acoustic memory means that the deaf have to learn a language without listening to it and this often makes it difficult for them to understand and produce spoken and written language. As far as reading comprehension is concerned, it can be defined as the convergence of two components: decoding and linguistic comprehension [1], where decoding refers to 'efficient word recognition, and linguistic comprehension is "the ability to take semantic information at the word level and derive sentence and discourse interpretations" [2].

Consequently, reading difficulties can be due to decoding problems, comprehension problems or to the interaction between the two aspects. Decoding and linguistic comprehension are both necessary to read but depend on different skills, in particular decoding skill is closely related to phonological skill while linguistic comprehension depends upon general language skill [3].

The literature on this topic has shown that the reading comprehension skill of deaf children is very low; specific research on reading skills of deaf students between 8 and 18 years of age detected a leveling off at the third – fourth grade [4,5,6,7,8,9].

The ability to identify the words is one of the main aspects determining reading comprehension. Nevertheless deaf children cannot access the spoken language and phonetic information, foreclosing in this way the identification of words based on phonemic skills [10,11].

Since in our society information is conveyed by means of the acoustic channel or the written language which present difficulties for deaf people, it is evident that a wide gap exist between people with normal hearing and the deaf. Of course, the difficulties of text comprehension persist in online reading too.

During the last few years, the new communication technologies have opened up many possibilities for the deaf to improve their life style.

The Communications Committee of the European Union states that “ICT plays a key role in the daily activities of any person. Whether it is employment, education, leisure, or at home, ICT is present in the form of computers, telephones and TV equipment. With the development of technology and the convergence of new platforms, these devices will play an increasing role in our daily activities. Many new and complex services are being offered through them. Access for people with disabilities to the Information Society is crucial for their integration and participation and their development as citizens with equal rights”.

2 Deaf and Technology: The System Gym2Learn

Technology can effectively help people with disabilities to be fully integrated into social life and there is a close link between machines and human wellbeing.

In particular, with regards to the reading skills of deaf students, various technological tools have been developed which aim to meet their specific needs.

However, the development of technology for deaf people is mainly focused on improving their ability to communicate with each other, on enhancing their linguistic knowledge and, at school level, on improving their competences and skills, but it does not take into consideration aspects outside the linguistic sphere.

Our work, instead, focuses on specific aspects of text comprehension and autonomous learning, using both traditional and technological educational tools.

In this framework, at the Institute for Educational Technology of the Italian CNR, we have developed Gym2Learn [12], a system to support online learning, based on a new annotation tool on web pages, which is easy for deaf people to use as well and has some features particularly suited to their needs.

Gym2Learn is a Mozilla Firefox browser add-on, extending the standard functionalities and allowing students to apply web annotations. The innovative aspect of the system is the process of creating notes linked to the application of cognitive strategies for text comprehension (recalling previous knowledge: formulating hypotheses and verifying them during surfing: asking and answering questions to verify comprehension: identifying important parts of the text). The system first trains the students in these strategies and then supports them in applying the strategies during online study and surfing.

The training area has been developed as a web site in which students do hyper-textual exercises and learn the strategies indicated above, becoming familiar with their main characteristics, the procedures for applying them and the context in which they are most effective.

The phase of web surfing, called the execution phase, offers students the opportunity to apply the strategies learned during the training phase. The student, through a simple graphical interface, can add his notes on web pages whose structure is isomorphic to the strategy selected. For example, when the student posts a note related to the strategy of "previous knowledge", the system asks him to describe the content he already knows and which he wants to link to an element of the page selected for the annotation.

The overall purpose of the annotation procedure is to give more importance to the student's cognitive and metacognitive processes than to the content of the annotation.

The activation of cognitive and metacognitive processes, in fact, assumes a key role in the processes of elaboration of the contents and has a positive influence on the student's performance.

The first release of the Gym2Learn system provided only a textual channel in the annotation mechanism.

After the first testing session with deaf students in a secondary school in Palermo, the results of which are soon to be published, it was decided to extend the functionalities of the system and include a new annotation mechanism with audio and video channels. This new multimodal approach makes it possible to target students with disabilities, providing them with a flexible system which caters for different learning styles and allows students to customize their learning activities according to their cognitive potential.

From a technical point of view, Gym2Learn is based on a client-server architecture consisting of a server for storing notes, a streaming server for managing audio and video resources and a client extension of the Firefox browser to access the system functionalities.

The annotations are modeled as a class of metadata. For the implementation of this architecture, the system adopts Annotea and extends the RDF Schema with the introduction of a specific namespace characterized by 4 types of annotation corresponding to the comprehension strategies implemented.

Each cognitive strategy is a type of annotation modeled according to an RDF schema in order to store properties about the strategy. This connection between comprehension strategy and type of annotation is implemented by extending the RDF schema with the metadata related to the properties of each comprehension strategy. In this way the action of creating an annotation corresponds to a specific comprehension strategy exercise that the student activates by selecting some texts on the document and using the pop-up menu to trigger the specific function.

In particular, in the case of deaf subjects, the use of the Gym2Learn video channel represents a new form of web annotation that can be defined as "Lis Annotation for Web Learning" (LAWL). In this approach, the deaf student can take notes using the linguistic channel that he likes best, by means of a video recording in LIS that corresponds to a specific cognitive strategy, in which he has been trained and supported by the ICT tool.

Gym2Learn can help students become autonomous during study and free online consultation, providing strategies that enable the user to make the most of his own capabilities and the various opportunities on offer on the web.

3 A Case Study

3.1 The Profile

Emilia is a 17 year old student with a profound bilateral neurosensorial hearing deficiency; her parents are deaf and they use the Italian Sign Language (ISL) to communicate. Emilia attends the fourth class of secondary school supported by a Personal Communication Assistant (PCA) and she speaks ISL as L1. Her scholastic performance is adequate and her favourite subjects are maths and art history.

She has the typical difficulties of hearing impaired (HI) students and in particular she shows problems in decoding, comprehension and in oral and written communication. Nevertheless, teachers did not schedule an individual curriculum path for Emilia that has enriched the planned learning goals with good results.

As far as the social area is concerned, Emilia is well integrated in the class, and has high motivation and a positive attitude to socialization. She uses the different accommodations easily and she likes the Internet very much, especially for chatting with other deaf friends. She has no difficulty in working autonomously, although she prefers asking her sign language interpreter to help her communicate with teachers and classmates because this is simpler and more convenient.

3.2 The Work Plan

The experimental phase was subdivided into ten sessions, one per week, each lasting ninety minutes; a sign language interpreter helped Emilia in carrying out the activities.

The first session aimed to familiarize her with the research team and the objectives of the study. For this purpose the team interviewed Emilia to find out about her study methods and specific learning habits and difficulties; she also completed two different questionnaires as described below.

During the second session, the researchers showed Emilia the Gym2Learn system and explained its theoretical background, focusing her attention on the importance of metacognitive strategies for learning both in traditional and online environments.

During this phase Emilia became familiar with the system and she was also supported in understanding how to use it correctly.

Then, three sessions focused on training in metacognitive strategies supported by Gym2Learn. As described above, in this area Emilia did some exercises to learn, directly online, the implemented metacognitive strategies and the ways for applying them effectively.

Afterwards, Emilia was engaged in three executive sessions during which she used the Gym2Learn functionalities to apply the strategies learned in the previous training phase for studying curricular topics. At the beginning of the activity, Emilia accessed a web page containing the task, the web resources to use and the instructions for producing a final piece of work which was later used to evaluate the activity.

The topic of study was global warming, including the causes, effects and solutions to this problem. During the navigation phase she could surf both suggested websites and websites of her own choosing, deciding how and when to adopt one or more of the text comprehension strategies.

The last session aimed to post test and evaluate whether the system had been useful in improving Emilia's online reading comprehension and metacognitive awareness in the use of cognitive strategies.

3.3 The Measure of Metacognitive Skills

The metacognitive skills were measured by the questionnaire Prova di Metacompreensione (PM) designed by Pazzaglia [10] to assess the metacognitive factors activated during reading and comprehension of a text. It consists of 34 questions which focus on the following areas (Tab. 1):

Table 1. Number of items in each area of the PM questionnaire

Area	No. Items
Knowledge of reading aims	8
Comprehension Check	8
Knowledge of strategies	8
Sensitivity to text	8

Most items have multiple choice answers and the student is asked to express her opinions about the reading aims, to identify the most effective strategy to carry out reading tasks, to judge the complexity of the text and the type of literary genre to which it belongs. In this study, we used the total score of the PM scale.

3.4 Self-evaluation of Metacognitive Skills

The total score of the MSI questionnaire was calculated to measure the increase of self-evaluation in metacognitive skills.

The MSI consists of 13 items subdivided into four areas (recalling previous knowledge; formulating hypotheses and verifying them during surfing; asking and answering questions to verify comprehension; identifying important parts of the text). For each item, the student must select one of four possible answers to best describe the most useful comprehension strategy.

4 Results

The results of the tests showed that the student was able to evaluate both positively and negatively the task she was performing but that she was unaware of the control and evaluation mechanisms she was using. Moreover, Emilia has sufficient command

of reading comprehension skills but is hampered in applying them by her disability. It should be pointed out that while Emilia was performing the tests she made some mistakes which can be explained with reference to the context of deafness as a specific life situation that also influences learning. In the MSI test, item number 4, "For good reading comprehension, it is a good idea:

- To read the text again to make sure that all the words are pronounced correctly.
- To ask for help with difficult words.
- To copy all the words in bold and italic.
- To evaluate if I have answered correctly."

Emilia responded "To ask for help with difficult words", although the correct answer was "To verify whether I have responded correctly to the questions"; in the same way, for item 7 where the question was the same, she responded: "To Ignore/skip words I do not know" while the correct answer was "To formulate hypotheses to guide my reading". These responses must be considered in the context of a deaf person who has difficulty understanding the meaning of a written text due to poor command of oral language because teaching strategies are focused on morphology and on the orthoepical emission of speech rather than on the meaning of words.

During the training phase, Emilia experienced difficulties in understanding descriptions, assignments, and evaluation tasks due to limitations in her vocabulary. In the evaluation it was therefore necessary to separate the accuracy of the task performed from the quality of the text produced. In fact difficulty in developing inferences, detecting inconsistencies, understanding differences between apparently similar information and interpreting metaphorical messages or idioms sometimes led the student to carry out the task inadequately from the point of view of content.

In performing the tasks, the student used the annotation system Gym2Learn correctly and applied the strategies she had learned, thus showing she understood their relevance and the functionalities of the system.

When she was using the system it was necessary to separate the evaluation of the task performance from the quality of the text produced. The notes she added, while containing errors of content and form, were correct with respect to the strategies she used. In particular Emilia preferred using the strategy "Asking questions".

The functionality provided by the system for creating video annotation meant that the supporting activity of the LIS interpreter was marginalized and favored the self-regulated study of contents.

Emilia only used the video annotation a few times during the experimental tests because she seemed a little embarrassed to be videotaped while taking notes. This difficulty decreased over time, as she was able to use the system at home and did not require the assistance of the interpreter.

A comparison between the two MSI tests revealed that the student modified her initial responses only in a few cases; in fact items 4 and 7, before mentioned, were not changed, showing that the student continues to use the dependent strategies for meta-comprehension.

It is significant that, compared to the pre-test, with respect to item 8, Emilia said that "For a better understanding of the text, it is a good idea: to copy the graphics and

tables that I find interesting."This answer seems to prove that the subject has been influenced by the use of the computer and the system. In the pre-test Emilia answered correctly, referring to the ability, to "ask myself questions". In the post-test, however, she seems to have responded as a result of what she had learned using the computer, or rather that you can copy diagrams and charts in order to better understand a topic.

5 Discussion and Conclusions

As a result of this brief experience of an ICT application with an HI student, it is important to discuss some of the qualitative outcomes. First, the post-test results highlight that at the end of the experiment E. had acquired greater autonomy when she read texts online. Thanks to the annotation functionalities of the Gym2Learn system, E. seems to be more confident and reflexive even when texts are difficult to comprehend.

She reported enthusiastically that by using the system she could search for and collect online information more easily and that her level of awareness when she used cognitive strategies was higher. In particular, of all the cognitive strategies implemented in the system to support reading comprehension, she favored the predicting and planning ones; in her opinion these strategies are very useful for achieving a specific goal of comprehension. At the same time, she discussed with researchers the importance of images and simulation because they are more stimulating and intuitive and, in particular, she was very surprised when a sign language interpreter appeared to accomplish the goal of assisting her in understanding texts she read. Her level of attention and concentration seems to have improved, thus reducing the support of the PCA.

It is important to note that, since the written production is influenced by the sign language interpreter, the errors made in adding notes must be contextualized and are not therefore to be attributed to cognitive deficits.

Although we are aware of the limited scope of this study, it seems reasonable to conclude that the use of Gym2Learn with deaf students can facilitate the development of personal autonomy and motivation to study, and promote the monitoring of strategies to apply in different tasks and in self-evaluation. The educational needs of deaf students cannot be separated from an increase in their ability to play an active role in studying and processing content, as the need is not just "to know", but also to reinforce self-esteem and a positive perception of individual limits.

Besides, it was expected or assumed that through repeated practice E. would learn these skills better and apply them to the new texts she read. Moreover, repeated practice of using the skills would enable her to internalize them, so they would become part of her reading repertoire.

Today, many of the technologies that are becoming increasingly available in the classroom are suitable for adaptation for hearing impaired (HI) students. Additionally, many newer technologies have become accessible and are provided through the school and these can enhance an HI student's educational program.

The system described above could be extended and made available to students with special needs so that these students are no longer isolated within the classroom, providing an inclusive environment in which education takes into account the specific needs of all the students.

References

1. Wauters, L.N., Van Bon, W.H.J., Tellings, A.: Reading comprehension of Dutch deaf children. *Reading and Writing* 19, 49–76 (2006)
2. Hoover, W.A., Gough, P.B.: The simple view of reading. *Reading & Writing* 2, 127–160 (1990)
3. Hilzensauer, M.: Information Technology for Deaf People. In: Ichalkaranje, N., Ichalkaranje, A., Jain, L.C. (eds.) *SCI*, vol. 19, pp. 183–206. Springer, Heidelberg (2006)
4. Furth, H.G.: A comparison of reading test norms of deaf and hearing children. *American Annals of the Deaf* 111, 461–462 (1966)
5. Conrad, R.: *The deaf schoolchild*. Harper and Row, London (1979)
6. Allen, T.E.: Patterns of academic achievement among hearing impaired students: 1974 and 1983. In: Schildroth, A.N., Karchmer, M.A. (eds.) *Deaf children in America*, pp. 16–206. College Hill Press, San Diego (1986)
7. Holt, J.A.: Stanford Achievement Test – 8th edition: Reading comprehension subgroup results. *American Annals of the Deaf* 138, 172–175 (1993)
8. Holt, J.A., Traxler, C.B., Allen, T.E.: *Interpreting the scores: A user's guide to the 9th edition Stanford Achievement Test for educators of deaf and hard-of-hearing students*. Gallaudet University, Washington (1996)
9. Traxler, C.B.: The Stanford Achievement Test, 9th Edition: National norming and performance standards for deaf and hard-of-hearing students. *Journal of Deaf Studies and Deaf Education* 5, 337–348 (2000)
10. Adams, M.J.: *Beginning to read: Thinking and learning about print*. Massachusetts Institute of Technology, Cambridge (1990)
11. Bradley, L., Bryant, P.E.: Categorizing sounds and learning to read: A causal connection. *Nature* 301, 419–421 (1983)
12. Chiazzese, G., Merlo, G., Seta, L., Ottaviano, S., Chifari, A., Allegra, M., Todaro, G., Ciulla, G.: *Gym2Learn – a Web Annotation System for text Comprehension*. *Learning Technology* 10, 11–13 (2008)

Extraction of Displayed Objects Corresponding to Demonstrative Words for Use in Remote Transcription

Yoshinori Takeuchi¹, Hajime Ohta², Noboru Ohnishi²,
Daisuke Wakatsuki³, and Hiroki Minagawa³

¹ Information and Communications Headquarters, Nagoya University,
Furo-cho, Chikusa-ku, Nagoya, 464-8603 Japan

takeuchi@nagoya-u.jp

² Graduate School of Information Science, Nagoya University,
Furo-cho, Chikusa-ku, Nagoya, 464-8603 Japan

{oota,ohnishi}@ohnishi.m.is.nagoya-u.ac.jp

³ Tsukuba University of Technology, Japan

{waka,minagawa}@a.tsukuba-tech.ac.jp

Abstract. A previously proposed system for extracting target objects displayed during lectures by using demonstrative words and phrases and pointing gestures has now been evaluated. The system identifies pointing gestures by analyzing the trajectory of the stick pointer and extracts the objects to which the speaker points. The extracted objects are displayed on the transcriber’s monitor at a remote location, thereby helping the transcriber to translate the demonstrative word or phrase into a short description of the object. Testing using video of an actual lecture showed that the system had a recall rate of 85.7% and precision of 84.8%. Testing using two extracted scenes showed that transcribers replaced significantly more demonstrative words with short descriptions of the target objects when the extracted objects were displayed on the transcriber’s screen. A transcriber using this system can thus transcribe speech more easily and produce more meaningful transcriptions for hearing-impaired listeners.

1 Introduction

Hearing-impaired students need complementary technologies that aid their speech understanding in lecture. Examples of such technologies are sign language interpretation, note-taking assistance, and PC captioning, which is already in use at several universities in Japan. A technology currently being developed for use at the university level is remote speech transcription. Kato *et al.* investigated the information required by a remote transcriber and ways to display that information on the transcriber’s monitor[1]. In the approach they developed, content keywords are displayed on the monitor to aid the transcriber. Miyoshi *et al.* are developing a remote real-time captioning system[2] that sends audio-visual signals from a classroom to a remote location where captionists type in real time. The system uses the “re-speak” method in which a captionist listens

to the instructor and then repeats what the instructor said[3]. The re-spoken speech is clear and distinct, so the recognition rate is high in the subsequent automatic speech recognition. Another technology that has captured much interest is real-time speech recognition. Wald and Bain, for example, are developing an automatic speech recognition system for universal access to communication and learning[4].

During a lecture, a university instructor will often project slides on a screen and point to different objects on them to clarify the explanation. When the instructor indicates a particular object or group of objects, he or she will typically utter a demonstrative word or phrase such as “this” and “these lines.” Transcribing these demonstrative words and phrases remotely and then displaying them on a screen in the classroom can take several seconds (~ 10 s). Since the instructor will likely finish pointing to the object before the demonstrative word or phrase is displayed, the students may have trouble connecting the subsequent explanation to the object that was indicated. Transcribers are typically aware of this problem and thus generally try to translate the demonstrative word or phrase into a short description of the object, which can be a difficult task, particularly if the transcriber did not see the target object. Previous research on identifying the target object includes analyzing the relationship between the pointing gesture and the motion of the stick pointer[5] and analyzing the relationships among the location and locus of the pointer end, the type of object, the position and orientation of the instructor, and the pointing gesture using sensors to detect the position and direction of the pointer and the position and orientation of the instructor[6]. None of the previous research included consideration of the demonstrative words or phrases uttered.

The system we are developing identifies the objects pointed to by the instructor along with the corresponding demonstrative words and phrases[7,8]. The identified objects are extracted and displayed on the transcriber’s monitor to help in translating the word or phrase into a short description of the object. We have now evaluated our system using two experiments, one to evaluate object extraction performance and one to evaluate the usefulness of displaying the extracted objects on the transcriber’s monitor. They showed that the system has high object extraction performance and that it significantly improves transcriber performance.

2 Identifying Target Objects

2.1 Recorded Audio and Visual Signals

With our system, a microphone is placed on the instructor’s lapel and a high-definition video camera is placed in the back of the lecture room (Figure 1). The sound signals are captured at a 16-bit/48-kHz sampling frequency, and the video signals have a 1440×1080 pixel size. The frame rate of the video signal is down-sampled at 10 Hz.

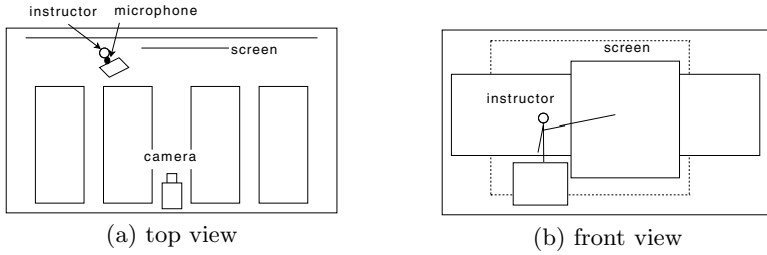


Fig. 1. System configuration in lecture room: (a) top view, (b) front view; dotted rectangle in (b) indicates image capture area

Table 1. Trajectory classifications

Category	Conditions
Point	More than half the points are extracted, and all the points are inside a small square.
Line	More than half the points are extracted, and all the points fall on a narrow line.
Other	More than half the points are extracted, but neither the point nor line conditions are met.
Ellipse	More than half the points are extracted, the point and line conditions are not met, and the trajectory surrounds an object.
Indeterminate	Less than half the points are extracted.

2.2 Extracting Pointing Gestures

The pointing gestures are extracted by focusing on the trajectory of the end of the stick pointer. The trajectory can generally be classified as a point, line, or ellipse. “Point” is when the instructor points to a specific location, “line” is when the instructor traces a line, and “ellipse” is when the instructor traces a curve around an object.

The system extracts the end of the pointer from five sequential frames and classifies the trajectory as point, line, other, ellipse, or indeterminate, as shown in Table 1.

If the trajectory is classified as a point, line, other, or circle, the system concludes that the instructor pointed to an object. It then displays the target object and the corresponding demonstrative word or phrase on the transcriber’s monitor to help him or her translate the word or phrase into a short description of the object.

2.3 Detecting Demonstrative Phrases

The demonstrative phrases are detected using speech recognition technology, specifically the Julius speech recognition engine 3.5. Investigation of audio signals recorded for an instructor revealed that the demonstrative phrases tended

to be spoken after a pause or a pause followed by a filler. We thus modified the probability settings of the language model so that it would recognize more demonstrative phrases in these situations. We also tuned the beam width, the language model weight, and the word insertion penalty. With these modifications, the system had a recall rate of the demonstrative word recognition of 91% and a precision of 71%.

2.4 Combining Pointing Gesture with Demonstrative Word or Phrase

The pointing gesture is combined with the demonstrative word or phrase to extract the target object. Investigation of the relationship between pointing gestures and demonstrative words and phrases using recorded video showed that the utterance of a demonstrative word or phrase occurs between 1.5 seconds before the start time and the end time of the pointing gesture. The formula for combining the pointing gesture with the demonstrative word or phrase is

$$t_{gs} - 1.5 < t_{ss} < t_{ge}, \quad (1)$$

where $t_{gs}(s)$ and $t_{ge}(s)$ are the start and end times of the gesture, respectively, and $t_{ss}(s)$ is the start time of the utterance of the demonstrative word or phrase.

2.5 Extracting Object(s) Corresponding to Demonstrative Word or Phrase

After combining the pointing gesture with the demonstrative word or phrase, the system extracts the corresponding object(s) using the following rules.

- If the trajectory is classified as “point,” extract the object nearest the point.
- If the trajectory is classified as “line,” extract all objects slightly above the line.
- If the trajectory is classified as “other,” extract all objects along the trajectory.
- If the trajectory is classified as “ellipse,” extract all objects surrounded by the trajectory.

The system extracts an image of the target object(s) and draws the trajectory of the pointer on the image, which is displayed on the transcriber’s monitor. Figure 2 shows an example of the system in use. The instructor had said, “These are $x(n)$, which are discrete-time signals,” while pointing to the bottom-left of the slide and moving the pointer in a circular manner. The system extracted this part of the slide and displayed it as shown at the bottom left. Subsequently, the instructor said, “These are z transformations of $x(n)$ ” while pointing at the equation with the pointer and not moving it around. The system extracted this equation and displayed it next to the previously extracted part.

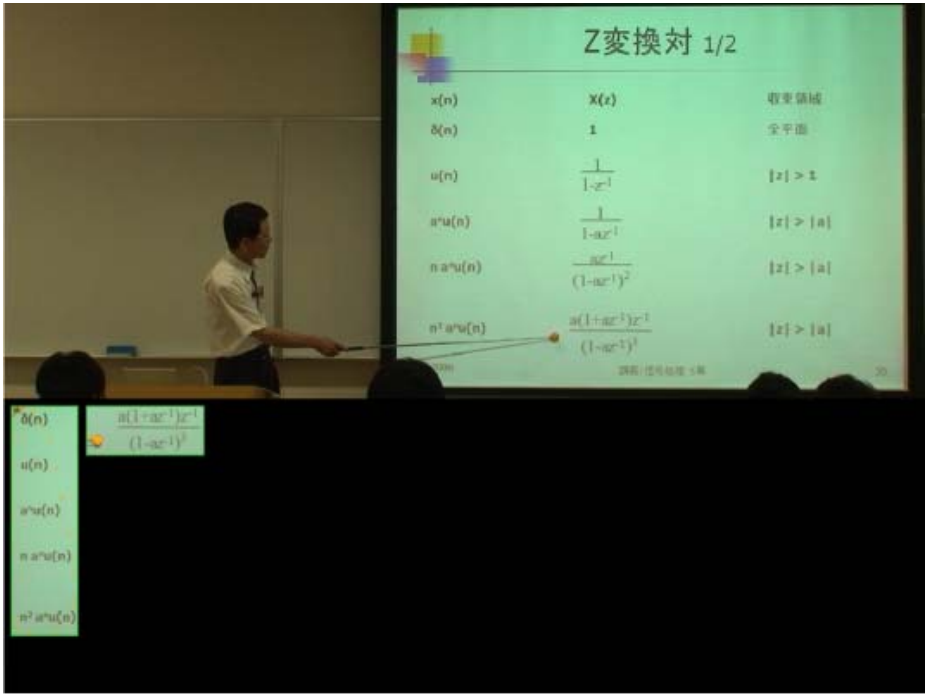


Fig. 2. Extraction of target objects: top part is image captured by camera and bottom part is image extracted for transcriber

3 Evaluation

3.1 Detection Rate

We conducted an experiment using audio-visual signals recorded during a 65-min lecture in which there were 189 pointing gestures with corresponding demonstrative words and phrases.

The result of the speech recognition is used as the time trigger in Eq.1 for object extraction. The t_{ss} is the true time detected by a person.

The system extracted 191 target objects from the video recording of the lecture. Table 2 shows the results.

3.2 Evaluation by Transcribers

We also conducted an experiment in which transcribers evaluated the usefulness of the object extraction and display. We showed the transcribers clipped lecture scenes containing brief pointing gestures, each with a corresponding demonstrative word or phrase. They were told to transcribe the words spoken and to replace the demonstrative words and phrases with a short description of the target objects. We showed them two scenes. The details of the scenes are summarized in

Table 2. Results for target object extraction

no. of pointing gestures with demonstrative words or phrases	189
no. of extracted objects	191
no. of correct extractions	162
no. of incorrect extractions	27
no. of excess extractions	2
recall rate	85.7% (162/189)
precision	84.8% (162/191)

Table 3. Two scenes used for transcriber evaluation

	scene A	scene B
duration	3m 40s	3m 55s
no. of words to be transcribed	15	17
no. of demonstrative words to be replaced	5	5

Table 4. Number of words transcribed: A and B with + means the transcriber was shown the extracted objects; A and B without + means the transcriber was not shown the extracted objects

	trans.	with	without
1st: scene A+	1	60% (9/15)	47% (8/17)
2nd: scene B	2	60% (9/15)	59%(10/17)
	3	47% (7/15)	47% (8/17)
1st: scene B	4	73%(11/15)	59%(10/17)
2nd: scene A+	5	60% (9/15)	53% (7/17)
	6	53% (8/15)	41% (5/17)
1st: scene A	7	65%(11/17)	33% (5/15)
2nd: scene B+	8	53% (9/17)	33% (5/15)
	9	59%(10/17)	40% (6/15)
1st: scene B+	10	88%(15/17)	60% (9/15)
2nd: scene A	11	47% (8/17)	47% (7/15)
	12	82%(14/17)	33% (5/15)

Table 3. As transcribers, we recruited 12 graduate students who could type very quickly and who would understand the content of the lecture.

We first gave the transcribers a transcript of the scenes and allowed them one minute to study it. We then grouped them into four groups. Each group viewed the scenes in a different combination of scene order and extracted object display (on or off), as shown in the first column of Table 4. The number of words transcribed is summarized in Table 4, and the number of demonstrative words replaced with a short description of the target object is summarized in Table 5.

Table 5. Number of demonstrative words replaced

	trans.	with	without
1st: scene A+	1	40%(2/5)	20%(1/5)
2nd: scene B	2	60%(3/5)	0%(0/5)
	3	40%(2/5)	40%(2/5)
1st: scene B	4	80%(4/5)	20%(1/5)
2nd: scene A+	5	60%(3/5)	20%(1/5)
	6	60%(3/5)	40%(2/5)
1st: scene A	7	60%(3/5)	40%(2/5)
2nd: scene B+	8	80%(4/5)	20%(1/5)
	9	40%(2/5)	0%(0/5)
1st: scene B+	10	80%(4/5)	60%(3/5)
2nd: scene A	11	60%(3/5)	40%(2/5)
	12	80%(4/5)	20%(1/5)

4 Discussion

The system achieved high performance (recall rate of 85.7% and precision of 84.8%) because it classifies the pointing gesture on the basis of the pointer's trajectory and then extracts the target object(s) on the basis of the classification.

The system did not extract 27 of the target objects. Some were among several objects indicated by a single one-point gesture combined with the word "these." Others were miss-combined with a demonstrative word or phrase. The extraction of two objects when the instructor was not pointing at anything was due to the pointer accidentally passing in front of the slide.

The transcriber evaluation showed that the transcribers could transcribe more words when watching both the lecture and the extracted objects than when watching the lecture only. They could also replace more demonstrative words with a short description of the object. A Wilcoxon signed-rank test showed that both results were significant at the 1% level. These results show that the proposed system can help transcribers replace demonstrative words and phrases with short descriptions of target objects.

5 Conclusion

A previously reported system extracts displayed objects corresponding to demonstrative words and phrases for use in remote transcription. The system identifies pointing gestures by analyzing the trajectory of the stick pointer and extracts the objects to which the speaker points. The extracted objects are displayed on the transcriber's monitor, thereby helping the transcriber to translate the demonstrative words and phrases into short descriptions of the objects.

We have now experimentally evaluated this system by using video of an actual lecture. The objects were extracted with a recall rate of 85.7% and a precision of 84.8%. We also experimentally evaluated its usefulness to transcribers by

using two extracted scenes. When the extracted objects were displayed on their screens, the transcribers replaced significantly more demonstrative words with short descriptions of the target objects. A transcriber using this system can thus transcribe speech more easily and produce more meaningful transcriptions for hearing-impaired listeners.

We are now working on reducing the incorrect extraction rate and the number of excess extractions. This involves improving the method used to estimate the pointing gesture, the method used to integrate the pointing gesture with the demonstrative word or phrase, and the method used to extract the target objects. We will then evaluate the system in a field test.

This research is partly supported by a KAKENHI Grant-in-Aid for Young Scientists (B) (21700577) from the Ministry of Education, Culture, Sports, Science and Technology of Japan and by the Telecommunications Advancement Foundation.

References

1. Kato, N., Kawano, S., Miyoshi, S., Nishioka, T., Murakami, H., Minagawa, H., Wakatsuki, D., Shirasawa, M., Ishihara, Y., Naito, I.: Subjective Evaluation of Displaying Keywords for Speech to Text Service Operators. *The Transactions of Human Interface Society* 9(2), 195–203 (2007) (in Japanese)
2. Miyoshi, S., Kawano, S., Nishioka, T., Kato, N., Shirasawa, M., Murakami, H., Minagawa, H., Ishihara, Y., Naito, I., Wakatsuki, D., Kuroki, H., Kobayashi, M.: A Basic Study on Supplementary Visual Information for Real-Time Captionists in the Lecture of Information Science. *IEICE Transactions on Information and Systems* (Japanese edition) J91-D(9), 2236–2246 (2008)
3. Miyoshi, S., Kuroki, H., Kawano, S., Shirasawa, M., Ishihara, Y., Kobayashi, M.: Support Technique for Real-Time Captionist to Use Speech Recognition Software. In: Miesenberger, K., Klaus, J., Zagler, W.L., Karshmer, A.I. (eds.) *ICCHP 2008*. LNCS, vol. 5105, pp. 647–650. Springer, Heidelberg (2008)
4. Wald, M., Bain, K.: Universal access to communication and learning: role of automatic speech recognition. *Universal Access in the Information Society* 6(4), 435–447 (2007)
5. Sakiyama, T., Mukunoki, M., Katsuo, I.: Detection of the Indicated Area with an Indication Stick. In: *Int. Conf. on Multimodal Interfaces*, pp. 480–487 (2000)
6. Marutani, T., Nishiguchi, S., Kakusho, K., Minoh, M.: Making a lecture content with deictic information about indicated objects in lecture materials. In: *AEARU Workshop on Network Education*, pp. 70–75 (2005)
7. Takeuchi, Y., Saito, K., Ito, A., Ohnishi, N., Iizuka, S., Nakajima, S.: Extracting Pointing Object with Demonstrative Speech Phrase for Remote Transcription in Lecture. In: Miesenberger, K., Klaus, J., Zagler, W.L., Karshmer, A.I. (eds.) *ICCHP 2008*. LNCS, vol. 5105, pp. 624–631. Springer, Heidelberg (2008)
8. Ito, A., Saito, K., Takeuchi, Y., Ohnishi, N., Iizuka, S., Nakajima, S.: A Study on Demonstrative Words Extraction in Instructor Utterance on Communication Support for Hearing Impaired Persons. *ibid*, pp. 632–639 (2008)

E-Scribe: Ubiquitous Real-Time Speech Transcription for the Hearing-Impaired

Zdenek Bumbalek, Jan Zelenka, and Lukas Kencl

R&D Centre for Mobile Applications (RDC)
Department of Telecommunications Engineering
Faculty of Electrical Engineering, Czech Technical University in Prague
Technicka 2, 166 27 Prague 6, Czech Republic
{bumbazde,zelenj2,lukas.kenc1}@fel.cvut.cz
<http://www.rdc.cz>

Abstract. Availability of real-time speech transcription anywhere, any-time, represents a potentially life-changing opportunity for the hearing-impaired to improve their communication capability. We present e-Scribe, a prototype web-based online centre for real-time speech transcription for the hearing-impaired, which provides ubiquitous access to speech transcription utilizing contemporary communication technologies.

Keywords: Hearing-Impaired, Real-Time Text, Voice over IP.

1 Introduction

Speech plays a basic role in communication between people, in education, in sharing ideas and in maintaining social contacts. The hearing-impaired have to challenge communications barriers in a mostly hearing-capable society. This barrier is especially serious when communicating with authorities or courts of law where inability to understand is a major obstacle. According to Czech law [18], hearing-impaired citizens have the right to choose a communication system that matches their needs. Today, this law is impossible to put in practice. One of the goals of the e-Scribe project, a joint activity with the Czech Union of the Deaf, is to remedy this situation. It may be overcome by using on-the-spot real-time speech transcription. Services of fast typists or sign-language interpreters may be employed. The concept of physically present transcribers is used for example in the project of the Czech Union of the Deaf: "Simultaneous transcription of speech" [17]. However, there is great shortage of educated transcribers, costs of these services are high and they are restricted to the particular location, therefore they can be offered to a limited amount of users only. An alternative is using Automated Speech Recognition systems (ASRs). Yet, ASRs are limited in recognition accuracy, especially of colloquial speech and of difficult national languages such as Czech.

The main contribution of the e-Scribe system is in designing a prototype centre for ubiquitous, cost-effective real-time speech transcription, first using human transcribers and later combining ASR with human error-correctors to

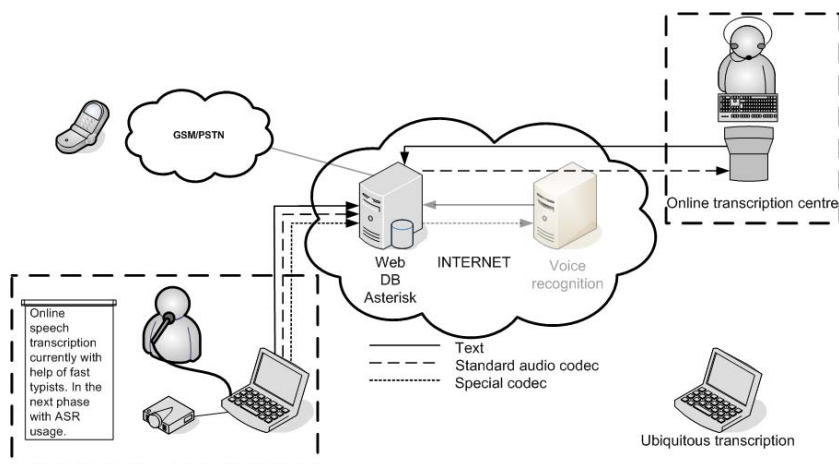


Fig. 1. e-Scribe Overview. Speakers's voice is transmitted via Voice over IP (VoIP) or telephony to the online transcription centre. The voice is transcribed by quick typist using MS Word[®], which is available via web form on the server. The transcribed speech is displayed on a projection screen at the conference using a web browser connected to the server. However, the transcription is available from any location with Internet connectivity.

automate the process of transcription (see Figure 1). The goal of the project is to set up an online transcription centre for the hearing-impaired, technically based on IP telephony and displaying the transcription online on a web page.

2 e-Scribe and Related Work

Using telephony for the deaf or hard-of-hearing has been possible by using text-phones known as Telephone Devices for the Deaf (TDDs) [7]. The disadvantage of TDDs is that both communicating users must have these textphones - yet only few hearing people own TDDs. For communicating with the hearing people, the Typetalk relay [14] could be used. The Short Messaging System is probably the greatest innovation in telephony for deaf people, but it can still not replace the real time flow of text, so to use SMS for longer transcription is impossible.

The topic of automated transcription can be explored from many points of view. An interesting enhancement in using ASR is the so-called captionist (transcriptionist) [3], [5] which is a person who re-speaks the speech that has to be transcribed, resulting in more intelligible and less noisy input for the ASR system. Experiments show [3] that re-speaking may significantly increase ASR efficiency. However, such a system needs at least one trained person for re-speaking the ASR input. ASR systems can be applied for example in instant messaging where the lower accuracy of transcription (about 10 % word error rate) [1], [5] has no fatal effect in understanding of sentence context. This approach is not

convenient when the transcription is supposed to be exploited in more public actions like lectures or technical meetings etc. where exact understanding is required. An interesting way to deal with transcription accuracy is complementing the ASR system with a person capable of error-correcting [4] of the ASR output. This can be done directly during the automated speech recognition. If it is not possible to do the correction immediately it can be done ex-post with the help of lecture audio record if necessary. Similar research in ASR for university lecture rooms is being carried out by the Liberated Learning Project (LLP) [15],[8]. The goal of LLP is enabling students with hearing impairment to access university lectures. In cooperation with LLP, IBM develops the ViaScribe [16] software that is specifically designed for real-time captioning.

The Villanova University Speech Transcriber (VUST) [6] was designed to improve the accessibility of computer science lectures using real-time speech recognition software. The VUST system is based on client-server architecture and consists of three major components: speech recognition, a dictionary enhancement tool, and a transcription distribution application. It is accessible locally via a wireless microphone in the classroom. In contrast, the e-Scribe is a ubiquitous system accessible from wide range of communication equipments such as laptops, PDAs, mobile or ordinary phones and real-time captioning is supported by skilled stenographers from the online transcription centre.

The idea of a ubiquitously available transcription tool was introduced in Scribe4Me: Evaluating a Mobile Sound Transcription Tool for the Deaf [2]. Upon pressing a button on a Scribe4Me PDA, the last 30 seconds of sound is uploaded, transcribed and sent back as a text message. The SMS communication presents a potential problem in interactivity due to the potentially significant delay. Additionally, the 30 seconds transcription is severely limiting. With today's mobile phones or PDAs, e-Scribe aspires to provide a real-time ubiquitous captioning via the phone web browser.

3 Implementation and Architecture

The currently operational e-Scribe solution is based on remote transcription carried out by quick typists. Voice is transmitted by VoIP telephony from the venue of the conference for the hearing-impaired (or another event) to the online centre, and then anywhere to the transcriber. Transcription is carried out by the specially trained typists who use a large list of abbreviations which are expanded into words or sentences. The typists are currently using Microsoft Word[®] software, due to historic, familiarity and performance reasons. Using MS Word[®] for textual input was one of their basic conditions for developing the application for transcription because for transcribers this environment is familiar and they are accustomed to specific layout, use of abbreviations and well-known behavior.

For the e-Scribe system MS Word[®] does not fulfill the requirements well due to the long delay when characters are transmitted. We are currently developing a web-based text editor simulating the behavior of MS Word[®] according to

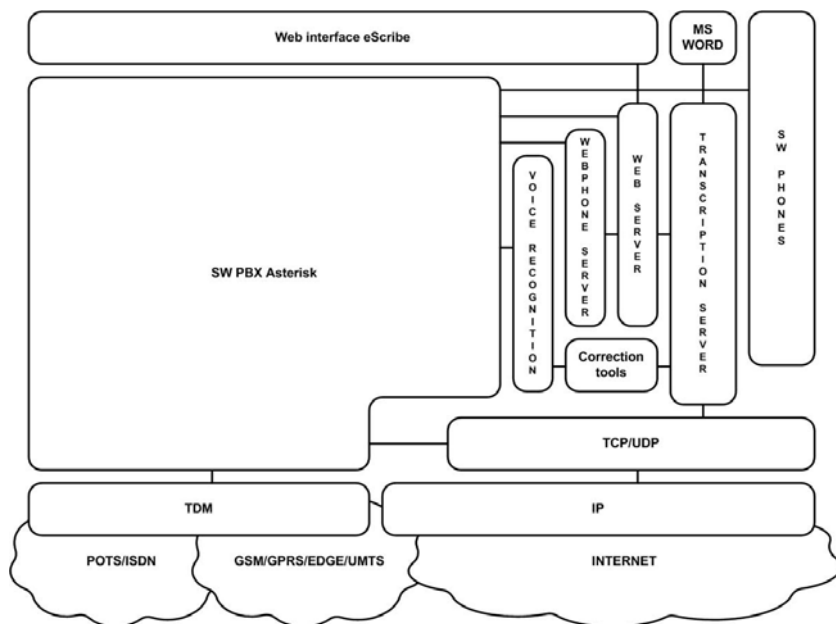


Fig. 2. e-Scribe Architecture. The core of the system, which represents the communication part of the project, is the software PBX Asterisk (Private Branch Exchange) [20]. The other block is the Transcription server which cooperates with the Apache web server [21]. Access to the system is possible from ordinary telephones, mobile phones but also from hardware SIP [11] telephones and software clients. The easiest way to access the system is using a web phone which makes the entire system available through the web browser interface, without necessity of any installation or configuration.

the requirements of transcribers. At this time we have designed a web application for testing purposes capable of text transmission, backward text editing, changeable text window, adjustable burst of characters to send, etc.. However, user transcriptions are currently still performed in MS Word[®].

The e-Scribe system is now successfully implemented and tested at lectures for the hearing-impaired students at CTU in Prague and it was also used at several public presentation events (e.g. the e-Scribe Press Conference in January 2010). The transcription system is available on the website [19] of the e-Scribe project and we are preparing to start official performance testing. The overall solution architecture is described in Fig. 2.

In IP telephony, Quality of Service is a frequent topic of discussion. Generally, in transcription systems it is very important to reach high-quality and reliable data flow to avoid information loss between speaker and transcriber. This demand requires not only a good internet connection, but also an errorless software chain and reliable hardware, if used. It is necessary to conduct testing in the form e.g. conversation tests which examine the whole communication chain and stability of software, hardware phones or of any other devices involved.

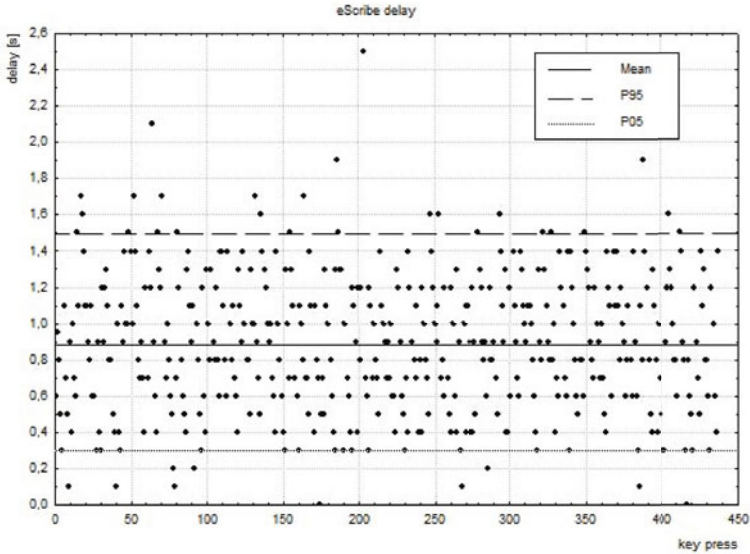


Fig. 3. Total delay from pressing a key to its displaying on a screen (450 values, mean and 5th and 95th percentile)

4 Performance Evaluation

In the process of transcription there are generally two main parameters for evaluation. The first parameter is the accuracy of the transcription. In the human-based transcription, this parameter depends on experience and dexterity of the typist. The second parameter is the delay between pressing a key and its displaying on a monitor or a projecting screen. To demonstrate the differences between situations where the typist is physically present and situations where he or she works remotely using the eScribe system, we prepared an experiment. We observed the total delay from pressing a key to its displaying on a screen. The delay was measured by evaluating the video record (25 frames/second) of keys pressing and displaying the transcription on the web page.

The experiment was carried out on a Windows-based laptop with MS Word[®] and a web browser. For video recording, the SW tool CamStudio [23] was used and the record was analyzed by the VirtualDub 1.9.8 [24]. The test was performed outside of the local network where the eScribe server is placed.

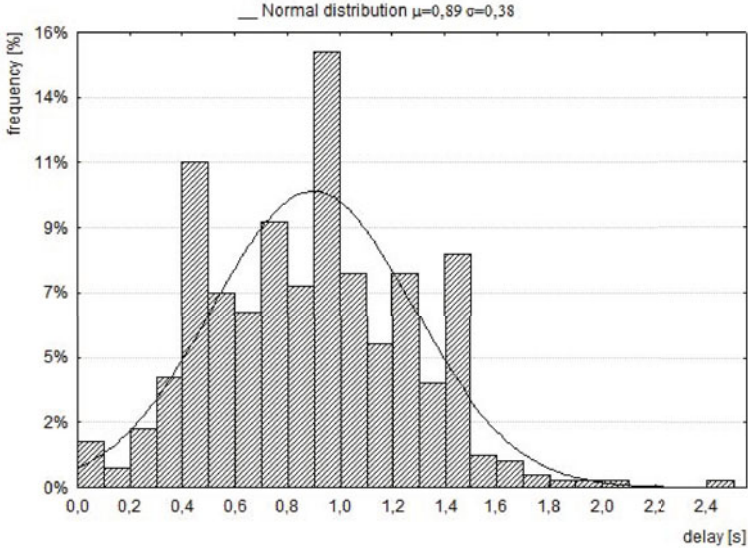
The total delay T_T consists of three partial delays:

$$T_T = T_W + T_N + T_D . \quad (1)$$

The T_W parameter represents the delay caused by MS Word[®] (time between pressing a key and sending it to the network) and due to native features of MS Word[®] the maximum is 1 second. The T_N parameter represents the Round Trip Time of the network and it was measured by the network debugging tool PING.

Table 1. Statistical evaluation

Total characters	Mean	Median	Modus	σ	σ^2	P_{05}	P_{95}
437	0.89	0.9	1	0.38	0.15	0.3	1.5
Network delay: Respons to ping request							
average delay			min		max		
22			10		329		

**Fig. 4.** Delay distribution

The parameter T_D represents delay caused by periodical refreshing of the web page with transcription. The refresh rate was set up to 550 ms in our experiment. The transcription was made by an averagely skilled typist, to whom an article of approximately 450 characters was dictated via VoIP telephony.

The average observed delay was 0.89 seconds with standard deviation 0.38 seconds. The delay distribution is depicted in Fig. 4 and the statistical parameters are summarized in Tab. 1. The delay variation is caused especially by the fact that the transcribed text is displayed in batches (corresponding to the amount of characters which the typist types during 1 second), but each character is compared with an individual key pressing.

In the presented experiment we have evaluated the architecture for online web-based transcription. Although according to the subjective experience of test users the total delay is still acceptable, it does not fulfill the requirements of the ITU-T.140 [25] recommendation for multimedia application text conversation specifying that a maximum buffering time may be 500 ms (300 ms recommended). It is evident that the crucial delay is caused by MS Word[®]. To reduce this, we are developing a web based editor simulating the MS Word[®] environment.

5 Future Work

The emphasis will be henceforth paid on ubiquitous web-based access. The current e-Scribe system is based on the standard client-server architecture using the open source Asterisk software [20] and the proprietary transcription application. The future goal is to propose the architecture and build a prototype of a universal web-based services, portable across various common mobile terminals, providing access to these assistive voice services to the handicapped anytime and anywhere, based on the emerging paradigm of Cloud Computing.

We will also integrate both voice and transcribed text into Asterisk using Real-Time Text (RTT) [13]. RTT is generally streaming text that is sent and received on a character-by-character basis and its benefits are summarized in [9]. In the Internet environment, RTT is represented by RFC 4103 [10] known as ToIP (Text over IP). The Session Initiation Protocol (SIP) [11] and the Real-Time Transport Protocol (RTP) [12] is used in ToIP for real-time text transmission. The ToIP framework is designed to be compatible with VoIP and Video over IP environments and could be implemented in current communication IP systems such as SW PBX Asterisk. With using VoIP and ToIP, a universal and standardized interface for interconnecting with the current and future ASR systems will be created, which is the main goal of the project. But because current speech-recognition software was found to be unsuitable for live transcription of speech (especially for national languages such as Czech), the transcription will be based on ASR software in co-operation with online human error correctors. This will make the transcription more cost-efficient and accessible to more hearing-impaired.

6 Conclusion

The widely accessible online transcription centre enables to provide services to much larger community of the hearing-impaired people. Remote transcription can offer this service with lower cost than if the typist were physically present at the conference or another event for the hearing-impaired people. This implies the amount of "transcribed actions" can be increased. Thanks to the online transcription, communication barriers will be minimized and cultural, educational, social or other events could be accessible for the hearing-impaired. Online transcription may also be used for communication with authorities or courts of law, where inability to understand is one of the most significant problems, or, thanks to its ubiquitous availability, eventually in day-to-day communication. Thanks to the e-Scribe system, communication barriers will be reduced and quality of life of the hearing-impaired people will improve dramatically.

Acknowledgments. The e-Scribe project is generously sponsored by Vodafone Foundation Czech Republic [22].

References

1. Wald, M., Bain, K.: Enhancing the Usability of Real-Time Speech Recognition Captioning Through Personalised Displays and Real-Time Multiple Speaker Editing and Annotation. In: Stephanidis, C. (ed.) HCI 2007, Part III. LNCS, vol. 4556, pp. 446–452. Springer, Heidelberg (2007)
2. Matthews, T., et al.: Scribe4Me: Evaluating a Mobile Sound Transcription Tool for the Delf. In: Dourish, P., Friday, A. (eds.) UbiComp 2006. LNCS, vol. 4206, pp. 159–176. Springer, Heidelberg (2006)
3. Miyoshi, S., Kuroki, H., Kawano, S., Shirasawa, M., Ishihara, Y., Kobayashi, M.: Support Technique for Real-Time Captionist to Use Speech Recognition Software. In: Miesenberger, K., Klaus, J., Zagler, W.L., Karshmer, A.I. (eds.) ICCHP 2008. LNCS, vol. 5105, pp. 647–650. Springer, Heidelberg (2008)
4. Wald, M.: Captioning for Deaf and Hard of Hearing People by Editing Automatic Speech Recognition in Real Time. In: Miesenberger, K., Klaus, J., Zagler, W.L., Karshmer, A.I. (eds.) ICCHP 2006. LNCS, vol. 4061, pp. 683–690. Springer, Heidelberg (2006)
5. Forman, I., Brunet, T., Luther, P., Wilson, A.: Using ASR for Transcription of Teleconferences in IM Systems. In: Universal Access in HCI, Part III, HCII 2009. LNCS, vol. 5616, pp. 521–529. Springer, Heidelberg (2009)
6. Kheir, R., Way, T.: Inclusion of Deaf Students in Computer Science Classes Using Real-Time Speech Transcription. In: Proceedings of the 12th Annual SIGCSE Conference on Innovation & Technology in computer Science Education, ITiCSE 2007, Dundee, Scotland, June 25–27, pp. 192–196. ACM, New York (2007)
7. Edwards, A.D.N.: Telephone access for deaf people (in) Home-Oriented Informatics and Telematics. In: Sloane, A. (ed.) Proceedings of the IFIP WG 9.3 HOIT 2005 Conference, pp. 235–244. Springer, New York (2005)
8. Bain, K., Basson, S., Wald, M.: Speech recognition in university classrooms. In: Procs. of the Fifth International ACM SIGCAPH Conference on Assistive Technologies, pp. 192–196. ACM Press, New York (2002)
9. Proposal R1 for Implementation of Real-Time Text Across Platforms, <http://trace.wisc.edu/docs/2008-RTT-Proposal/>
10. Hellstrm, G.: IETF RFC 4103: RTP Payload for Text Conversation, the Internet Society (2005)
11. Rosenberg, J.: IETF RFC 3251: SIP: Session Initiated Protocol, the Internet Society (2002)
12. Schulzrinne, H.: IETF RFC 3550 RTP: A Transport Protocol for Real-Time Applications, the Internet Society (2003)
13. <http://www.realtimetext.org>
14. <http://www.mid-typetalk.Org.uk>
15. The Liberated Learning Consortium, <http://www.liberatedlearning.com/>
16. IBM ViaScribe, http://www-03.ibm.com/able/accessibility_services/ViaScribe-accessible.pdf
17. Czech Union of Deaf, <http://www.cun.cz>
18. Czech law no. 155/1998 on comm. systems for the hearing-impaired, as amended by Act no.384/2008
19. <http://www.escribe.cz>

20. Meggelen, J., Madsen, L., Smith, J.: Asterisk: The Future of Telephony, 2nd edn. O'Reilly Media, Sebastopol (2007)
21. The Apache Software Foundation server, <http://www.apache.org/>
22. The Vodafone Foundation Czech Rep., <http://www.nadacevodafone.cz/>
23. <http://www.rendersoftware.com/products/camstudio/>
24. <http://www.virtualdub.org/>
25. ITU-T recommendation T.140: Protocol for multimedia application text conversation, ITU-T (1998)

SportSign: A Service to Make Sports News Accessible to Deaf Persons in Sign Languages

Achraf Othman, Oussama El Ghoul, and Mohamed Jemni

Research Laboratory of Technologies of Information and Communication
Ecole Supérieure des Sciences et Techniques de Tunis
5, Av. Taha Hussein, B.P. 56, Bab Mnara, 1008 Tunis, Tunisia
{achraf.othman,oussama.elhgoul}@utic.rnu.tn,
mohamed.jemni@fst.rnu.tn

Abstract. Sports are important in the life of deaf, as well as hearing persons, on physical, social and mental levels. However, despite that there exist many deaf sports organization in the world, the announcement of sports' results and actualities is still done by written or vocal languages. In this context, we propose a Web based application to edit and diffuse sports news in sign languages. The system uses the technology of avatar and it is based on WebSign kernel developed by our research Laboratory of Technologies of Information and Communication (UTIC). Information is broadcasted using MMS containing video on sign languages and also published in SportSign's Web site.

Keywords: Sport, Deaf, Sign Language.

1 Introduction

Deaf persons meet real problems to know sports results unless they are not looking to TV or not present on site as stadiums. Many deaf persons are interested to know news and sport results of their clubs or their national teams as well as important sport events as the Olympic Games and international tournaments of football. However, a big problem of accessibility to the information arises. Due to the large number of illiterate deaf [11], 80% of deaf person around the world in 2003 (accordingly to the World Federation of Deaf [12]), the access to the textual information is still difficult. In fact, medias presenting this kind of news in sign languages are very rare. It is the same situation concerning systems for translating in real-time textual sport news into sign languages. However, for some important competitions, like the football world cup or the Olympic Games, obtaining the results in real time (or in a short period of time) is important. Remark that there are some IT solutions proposing the diffusion of sport news in textual mode via SMS which are inaccessible for the majority of deaf persons.

In this context, we present in this paper a new system which allows the edition of sports results and actualities in sign languages. Results are diffused in a web site or via MMS to allow real time sign language news' broadcasting. The system is based on web technologies and it uses WebSign [5] service to translate textual content to sign languages.

This paper is organized as follows: the second section presents the importance of sports for deaf and the efforts done to consolidate sport culture in deaf community. The third section describes the platform SportSign allowing the elaboration of sports news in sign language using a friendly of use interface. In addition, it illustrates the architecture of the platform and how we can make and broadcast MMS containing sign language news. We conclude the paper by a conclusion.

2 State of the Art

During our survey of the literature, we have not found websites or others information systems dedicated to diffuse sporting events and news in sign languages. Although, there exist several websites for deaf sporting which enclose only textual contents. For this purpose, we emphasis in this section the importance of sport in deaf life and we present the state of the art related to existing technologies for diffusing actualities and sport news in sign languages.

2.1 Sport and Deaf

Thanks to the efforts of civil society, to media coverage, to political and public willing, considering how people with disabilities can effectively access to services becomes more and more popular and demand of social integration of people with disabilities is increased, including the sporting world. In this context, introducing sporting in the life of the deaf occurred initially during the 20th century. The deafness or hearing impairment is a full or partial inability to detect or perceive at least some frequencies of sound. During the last hundred of years, there is no specialized sporting for deaf. However, many sports are based essentially on sounds as signals. As deaf people have entered the sport world, various terms have evolved to represent their involvement as perceived by the able-bodied sport in the world.

On one level, deaf sport can defined as sport in which deaf players compete. On another level, deaf sport should be viewed from a cultural perspective. As noted above, there has been a sharp increase in the range of so-called DPOs [1] “Disability People’s Organizations” that have engaged in lobbying for the sport. These DPOs are too active, as an example the International Committee of Sports for the Deaf who organizes the Deaflympics, and presents a web portal that contains the results and actualities. However, for an illiterate deaf, it is impossible to read and follow results in absence of an expert who translates them in the sign language.

For this reasons, many techniques focus to find a solution that allows deaf persons to follow the latest results.

The next section presents some techniques developed to allow deaf persons to follow sports results and actualities. We present, in particular, two techniques present in the literature that make news and actualities more accessible. The first technique is the subtitles for video and audio content and the second is a video player embedded into web portals and containing a video sequence that translates text to sign language by an expert or deaf persons.

2.2 Using Subtitles in Video Contents

TV and video streaming on the web present a very important source of information and news. However, they are not accessible for deaf persons, particularly the results of sporting events.

In this context, a first technique allows deaf people to understand video sequences by adding subtitles. The subtitles must be synchronized with video and contain more information and coloration. For this reason, Microsoft developed the SAMI language [2]. A SAMI file can be created with a simple text editor. It has the extension “.smi” and the same name as the video sequence file. Then, to play video, we have just to put into Windows Media Player the video file and the SAMI file. The subtitles will be therefore synchronized with the video. Furthermore, another subtitling language has been proposed by the Italian Alessio Corticci based on SMIL language [3]. The SMIL file worked only in FLV player. Remark that, this solution makes video streaming more accessible and understandable for deaf who can read and write but not for illiterate deaf persons.

2.3 Video Based Solutions

The second solution used in some websites for deaf persons presents sport results in sign language via FLV player playing the sign language’s translation of textual content of actualities. The translation is made by an expert or by a deaf person. For example, the official website of UK deaf sport [4] presents some sports results through FLV player in the right section of the web page (see figure 1).

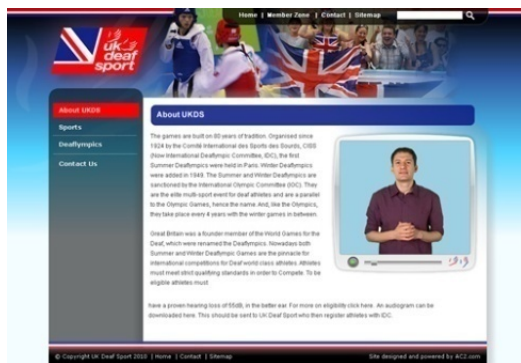


Fig. 1. UK Deaf Sport Official Website

In fact, this solution makes textual news and sports results accessible for deaf people but its inconvenient that it requires experts in sign language for recording the translation. In addition, there are too many websites for deaf sports which present the information in textual format due to the cost of the video based solution. For instance, the Deaflympics official website, does not contain any alternative in order to make textual content accessible for illiterate deaf person.

In order to provide solutions to the inconvenient mentioned above, and to make news and sports results accessible to all deaf persons, our solution is based mainly on avatar technology. This solution is described in next section.

2.4 Avatar Based Technology

Many websites translate their textual content to video sequences in sign languages. However, this technique has many limits. The first one is the difficulty and the cost of video storage and interpretation. The second is that any update on textual content needs to restore all video sequence. In addition, the transmission of video is a bandwidth consuming. As a new technology, many developers and researchers prefers to use avatar based approaches and virtual reality to cover the gaps of video systems.

The WebSign kernel [5] developed within our research laboratory UTIC is based on the avatar technology, allowing the interpretation of a text in natural language to sign language in real time. This interpretation is built through a dictionary of words stored in a database.

The creation of sign language words in the dictionary may be made in an incremental manner by users who propose signs. The words which make up the dictionary are validated by an expert in sign language. And as the sign language is not universal, the Websign toolkit offers a multi-community approach.

Technically, the WebSign kernel integrates a player of signs based on avatar interpretation which analysis the SML files (Sign Modeling Language) [6] and generates movements. The SML file is in fact an XML file that codes the movements of gestures of the avatar. WebSign is based on a Web-service architecture which allows remote service for sign language translation. This web service can be used by deaf accessible applications. For instance, we have opted to use this service by our application to generate signed contents of sport results.

3 SportSign Service Platform

In order to contribute in the consolidation of sport culture within deaf people community, we have developed a new sport portal, called SportSign, which offers actualities in sign language. SportSign is an application allowing deaf people to follow up the latest sports actualities and results via MMS messages and/or via the website. Our application offers mainly an environment to create, broadcast and consult various sports results in sign language animated by WebSign's avatar. These results can be published via the Web using FLV technology or via multimedia messages MMS for cellular phones.

3.1 Our Contribution

SportSign is a web portal which aims at publishing sports news using two main new technologies blended together (Web and MMS). The home page of the portal (see figure 2) contains an FLV player animating a welcome message in sign language. In addition, a navigation bar composed of a label and an image description for making the menus more understandable and accessible. It encloses also, a textual zone presenting the latest sporting results. In order to make textual menus more accessible for

deaf persons, a graphical area containing a set of images, representing the finger spelling of mouse pointed menus, will be shown.

All news are represented in sign language using FLV contents. Although, The SportSign portal requires that the guests of the web site register in order to watch the video sequences. The registration does not need too many information, just the name, the password and an email address. Such information can help us to make statistics about connecting users and to send them newsletters containing SportSign updates and events.



- ❶ FLV player
- ❷ Navigation Bar
- ❸ Latest sport results
- ❹ Online Sign translator
- ❺ Login Area
- ❻ Polls

Fig. 2. SportSign home page

3.2 Technical Architecture

Thanks to the evolution of web technologies, new views of web architecture have appeared, like web services [8]. The web becomes a net of services connected together via remote procedures. Following this view, we have built our system using principally Web-Services architecture. The proposed architecture is based on three main services: the first is a service for translating texts to sign language (WebSign Service), the second is dedicated to transform SML description of sign into video based signed sentences and the third is used to send MMS containing the generated video using telecom provider service and the MM7 protocol [9].

As shown in figure 3, the SportSign platform is based on distributed architecture which allows updating and modifying each module independently. In addition, this design allows the exploitation of each service in another project or other purposes.

As illustrated in the figure above, the information is published in two different ways: the first is the web and the second is the MMS. The two techniques need that the video content has to be available in the server. These contents are generated by the

administrator by means of an editor (described in the next section). When the administrator creates the news, SportSign invokes the remote procedure of WebSign service in order to retrieve an SML description of the corresponding sign. The next step consists to transform SML description into two video formats containing signed sentences (The FLV format, dedicated to the web, and the 3GP format, dedicated to the MMS). The generation of video based signed content from SML is a hard task and needs several minutes of computation. For this reason, it is more appropriate to be executed on a separated machine (as a web-service). Once the videos are generated, SportSign will automatically update the list of news in the website and broadcast MMS to the mobile phones of subscribed people.

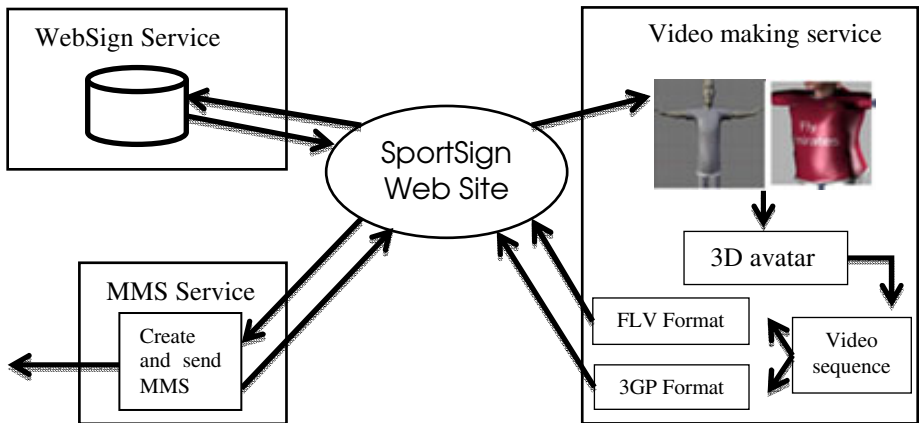


Fig. 3. Sport Sign architecture overview



Fig. 4. Sport results editor

3.3 News Editor

Sport news and results are built by the administrator thanks to an easy of use editor (see figure 4). Its interface is also a Web based technology and it is accessible only for the administrator. In the first step, the administrator has to choose the type of sport which concerns the actuality: football, basketball and so on. According to the selected type of sport the system displays the list of information that should be edited.

For example, if the selected sport is football, the editor displays some graphical components allowing the administrator to choose the two teams and to write the number of goals scored by each team. According to the result, the video generator service can make the avatar with the shirt of the winner team.

In addition, the editor allows writing a short description of the event which will be shown as textual information dedicated to literate deaf. Finally the administrator should specify if this event will be broadcasted by MMS to subscribed members or not.

When the administrator validates the written event, the system starts the process of video generation and publishes automatically a short video containing the game result in sign language.

4 Conclusion

In this paper, we presented a new application for monitoring and publishing of sports results based on WebSign kernel. The originality of this tool is the generation and dissemination in real time of sports results via the web and cell phones. And this contributes also in making the web and mobile phones more accessible to the deaf community. This platform has been tested via a telecom provider in Tunisia and it will be soon available online. As future perspective, we plan to extend this platform to cover different kind of news such as economy and news concerning deaf associations. In addition, we plan to add the notion of multi-communities, allowed by WebSign Kernel, because sign language is not universal.

References

1. Levermore, R., Beacom, A.: Sport and International Development (Global Culture and Sport). Palgrave Macmillan, Basingstoke (2009)
2. Understanding SAMI 1.0, <http://msdn.microsoft.com/en-us/library/ms971327.aspx>
3. Hoschka, P.: An introduction to the synchronized multimedia integration language. *IEEE Multimedia* 5(4), 84–88 (1998)
4. UK Deaf Sport Official Website, <http://www.ukdeafsport.org.uk/>
5. Jemni, M., Elghoul, O.: A System to Make Signs Using Collaborative Approach. In: Misesenberger, K., Klaus, J., Zagler, W.L., Karshmer, A.I. (eds.) ICCHP 2008. LNCS, vol. 5105, pp. 670–677. Springer, Heidelberg (2008)
6. Jemni, M., EL Ghoul, O.: Towards Web-Based automatic interpretation of written text to Sign Language. In: ICTA 2007, Hammamet, Tunisia, April 12-14 (2007)
7. Jemni, M., EL Ghoul, O., Yahia, N.B., Boulares, M.: Sign Language MMS To Make Cell Phones Accessible To Deaf. In: Hersh, M.A. (ed.) Conference & Workshop on Assistive Technologies for People with Vision & Hearing Impairments Assistive Technology for All Ages CVHI (2007)

8. Erl, T.: *Service-oriented architecture: a field guide to integrating XML and web services*. Prentice Hall PTR, Upper Saddle River (2004)
9. Guthery, S.B., Cronin, M.J.: *Developing MMS applications: multimedia messaging services for wireless networks*. McGraw-Hill Professional, New York (2003)
10. Huenerfauth, M., Hanson, V.L.: *Sign Language in the Interface: Access for Deaf Signers*. *The Universal Access Handbook*, 38.1–38.14 (2009)
11. Marschark, M., Harris, M.: *Success and failure in learning to read: The special case of deaf children*. *Reading comprehension difficulties: Processes and intervention*, 279–300 (1996)
12. World Federation of the Deaf (WFD), *Position Paper regarding the United Nations Convention on the Rights of People with Disabilities, Ad Hoc Committee on a Comprehensive and Integral International Convention on the Protection and Promotion of the Rights and Dignity of Persons with Disabilities* (2003)

Synote: Accessible and Assistive Technology Enhancing Learning for All Students

Mike Wald

School of Electronics and Computer Science,
University of Southampton, United Kingdom
M.Wald@soton.ac.uk

Abstract. Although manual transcription and captioning can increase the accessibility of multimedia for deaf students it is rarely provided in educational contexts in the UK due to the cost and shortage of highly skilled and trained stenographers. Speech recognition has the potential to reduce the cost and increase the availability of captioning if it could satisfy accuracy and readability requirements. This paper discusses how Synote, a web application for annotating and captioning multimedia, can enhance learning for all students and how, finding ways to improve the accuracy and readability of automatic captioning, can encourage its widespread adoption and so greatly benefit disabled students.

1 Introduction

Text transcriptions of the spoken word can benefit deaf people and also anyone who needs to review what has been said (e.g. at lectures, presentations, meetings etc.) The provision of synchronized text captions (subtitles) and images with audio and video enables all their different communication qualities and strengths to be available as appropriate for different contexts, content, tasks, learning styles, learning preferences and learning differences. For example, text can reduce the memory demands of spoken language; speech can better express subtle emotions; while images can communicate moods, relationships and complex information holistically. Deaf learners and non-native speakers may be particularly disadvantaged if multimedia involving speech is not captioned while visually impaired learners may be disadvantaged if accessible text or spoken descriptions of visual information is not available. Multimedia has become technically easier to create (e.g. recording lectures) but while users can easily bookmark, search, link to, or tag the **WHOLE** of a podcast or video recording available on the web they cannot easily find, or associate their notes or resources with, **PART** of that recording. As an analogy, users would clearly find a text book difficult to use if it had no contents page, index or page numbers and they couldn't insert a bookmark or annotation. Therefore the growing amount of knowledge available in multimedia formats has yet to achieve the level of interconnection and manipulation achieved for text documents via the World Wide Web and so realize the exciting opportunities for learning. This paper discusses how captioning has the potential to enhance learning for all students through making multimedia web resources (e.g. podcasts, vodcasts, etc.) easier to access, search, manage, and exploit for

learners, teachers and other users through supporting the creation of synchronised transcripts, notes, bookmarks, tags, links and images. The paper also argues that the availability of accurate automatic captioning using speech recognition (SR) will encourage the widespread adoption of this technology and so also greatly benefit disabled students. Research has confirmed the importance of captions for searching recordings and reading the transcripts and the value of also being able to personally annotate the recordings (e.g. bookmarks, notes and tags) and search these annotations [1], [2], [3], [4]. Speech recognition has been demonstrated to provide a cost-effective way of automatically creating accessible text captions and transcripts synchronised with audio and video and so allowing audio visual material to be manipulated through searching and browsing the text [5]. Although real-time captioning using phonetic keyboards can provide an accurate live transcription for deaf people, it is often not available because of the cost and shortage of highly skilled and trained stenographers [6]. Real-time SR is required for deaf and hard of hearing students and non native speakers whereas non real-time SR may often suffice for others and will provide more time for processing allowing for more accurate methods of SR to be used. Available SR systems (e.g. Dragon, ViaVoice [7]) can be adjusted to provide the fastest recognition (e.g. real-time) or to provide a slower but more accurate text output. It is possible therefore to provide fast but less accurate real-time recognition in a lecture and then later replace the transcript with one produced using slower but more accurate non real-time recognition. A study of the use of SR captioning in classrooms [8] showed that students with disabilities liked the fact they were not the only people to benefit from the technology as it drew the entire class into a collective learning experience and so making the recording and captioning of lectures standard procedure in universities would be of great benefit.

2 User Needs

User needs analysis has identified many benefits of annotating multimedia recordings in the ways proposed in this paper. It will for example enable learners to: search text transcripts, slides and notes and then replay recordings from that point; read captions to support learning style preference, deafness, or second language; read text descriptions of visual information (in videos and/or slides); use the transcript of lectures to support their understanding of text books and academic articles if they find the more colloquial style of transcribed text easier to follow than an academic written style; insert a bookmark in a recording so as to be able to continue later from where they left off; link to sections of recordings from other resources (e.g. documents, web pages etc.) or share these sections with others; tag and highlight sections of recordings/transcripts they don't understand fully so they can revisit them later for clarification; annotate recordings with notes and URLs of related resources (e.g. documents, websites etc.) at specific places in a recording to clarify issues and support revision; tag recordings using their own terms as a personal index. It will also enable teachers/lecturers to: index their recordings using syllabus topic tags; provide synchronized slides and text captions to accompany podcasts; provide text descriptions

for visually impaired students of visual information; identify topics needing further clarification from the pattern of learners' 'not understood' tags; provide feedback on learner-created recordings of presentations; ask learners to annotate recordings to provide evidence of their group contributions; analyse unstructured tags learners use (folksonomy) to help create structured tags (ontology); tag recordings with URLs of related resources (e.g. documents, websites etc.); link to and use sections of existing multimedia without having to edit the recording. While some proprietary existing systems can synchronise students' notes with teachers' presentations and recordings, they do not provide captioning (e.g. Tegrity [9]). Since no existing technology was found to satisfy all the identified user needs it was decided it was therefore necessary to develop Synote, a new web application that: Works with web multimedia and stores annotations separately in XML format; Synchronises captions, images, tags, links, notes and bookmarks; Enables users to add and search for annotations quickly and easily; Supports private or shared annotations; Is accessible to people with any disability.

3 Automatic Speech Recognition Captioning

Commercially available SR software (e.g. Dragon, ViaVoice) is unsuitable for transcription of speech in lectures as without the dictation of punctuation it produces a continuous unbroken stream of text that is very difficult to read or comprehend. Liberated Learning (LL) and IBM therefore developed ViaScribe [10] [11] as a SR application that automatically formats synchronized real-time text captions from speech with a visual indication of pauses. Detailed feedback from students, and lecturers [8] showed that this approach enhanced teaching and learning if the text transcription was reasonably accurate (e.g. <15% word error rate). Similar results have been reported by others using the Microsoft SR Engine [12] or Dragon [13]. To improve accuracy and understanding of SR captions an application RealTimeEdit was developed to enable corrections of SR captions to be made in real-time [14]. IBM has also recently developed a speaker independent recognition engine, Attila [15] that can that can be hosted on the web, interfaced with other applications and used to transcribe, edit and display recordings created in a wide range of multimedia file formats. The Attila speech recognition system has only available a US and not a UK English voice and language model and has no simple facility for adding vocabulary or training to individual users' voices and so typically gives word error rates between 15% - 30% for UK speakers using headset microphones. This however compares well with the National Institutes of Standards (NIST) Speech Group reported WER of 28% for individual head mounted microphones in lectures [16]. For speech recognition systems to be able to transcribe lectures more accurately in addition to having 'local' (e.g. UK) language and voice models they need to be designed for education rather than for dictation. For example they need to: Be speaker independent with a speaker dependent training facility; Be customisable for different subject domains; Cope with low quality speech signals and background noise; Recognise or ignore partial words (e.g. hesitations) or

‘fillers’ (e.g. ‘um’, ‘er’). Commercial rates for manually transcribing and synchronising a lecture recording are typically around £2/minute [17] (rates vary dependent on quality and quantity) and so it would cost about £90 for transcribing and synchronising a 45 minute class. For speech recognition to be used it must therefore cost less than this including licensing, server and maintenance costs. Manual correction of errors will also be required if 100% accuracy is to be achieved. One possible sustainable approach to obtaining accurate transcriptions could be to devise a system where students in the classes themselves corrected errors they found in the transcript, either voluntarily or through being paid or through being given academic credit.

4 Synote Annotation System

A multimedia annotation system called Synote [18] that meets the identified requirements has been developed (supported by the JISC [19]) and trialed (supported by Net4Voice [20]), providing multimedia recordings synchronized with transcripts, slides images and bookmarks (called ‘Synmarks’ and containing titles, tags, notes, and links) [21]. Synmarks can also be used to provide information about sounds or tone of voice or emotions or synchronised links to videos of sign language translations. Synote has been integrated with Attila to simplify automatic transcription. Figures 1 & 2 show examples of Synote’s interface while Figure 3 provides a system overview. When the recording is played the currently spoken words are shown highlighted in the transcript. Selecting a Synmark, transcript word or slide image moves the recording to the corresponding time. Synote also enables searching of the transcripts, Synmarks and PowerPoint slide titles, notes and text content. Users can also synchronise, or transcribe transcripts by hand if they don’t have access to suitable SR software that automatically transcribes and synchronises speech and text. Synote is currently being used by teachers and students in universities throughout the world.

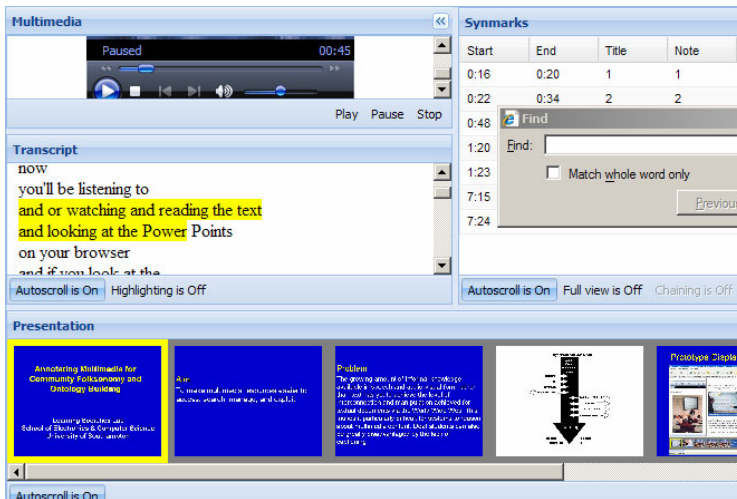


Fig. 1. Synote Interface

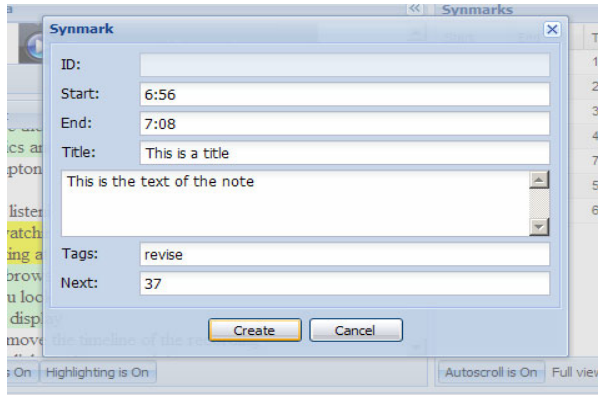


Fig. 2. Creation of Synmarks

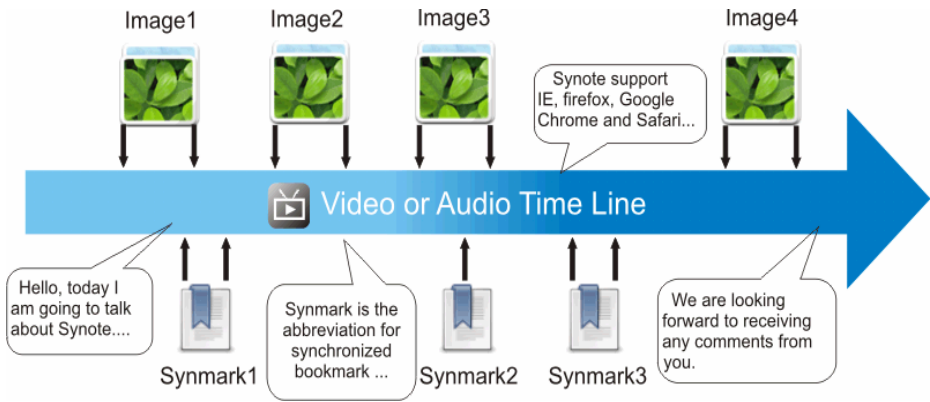


Fig. 3. System Overview

Since 2007 Dr Wald has used Synote with over 40 recordings of his lectures with synchronised transcripts and slides for his teaching of approximately 200 students on 5 undergraduate and postgraduate modules in The University of Southampton’s School of Electronics and Computer Science (ECS). At the time of writing there are over 500 recordings available on Synote (most with synchronised transcripts) for teachers and students to use for teaching and learning. These include guest lectures presented by leaders in their field. Questionnaires and interviews with students and teachers who have used Synote have shown that both students and teachers like the system and have found it enhances learning and teaching and would like more recordings of classes to be made available in this way. In the University of Southampton School of Electronics and Computer Science (ECS) 5 classes with approximately 200 students (two 1st year BSc classes, one 3rd year BSc class and two Masters Degree classes) were surveyed, with 101 students filling in questionnaires about their experience with Synote. The questionnaire results showed that Synote’s design to support a wide range of browsers was justified and appreciated with 54% using Internet

Explorer, 31% using FireFox, 12% using Google Chrome and 3% using Safari. The results also confirmed that Synote was easy to use as over 80% of the respondents didn't need to read or listen to the Synote guides and rated Synote as 4 or 5 on a 5 point scale of ease of use with the remaining percentage rating it as 3. The design of Synote to provide synchronised slides, video, audio and transcripts was also appreciated with over 80% respondents rating Synote as 4 or 5 for being useful overall as well as its slides, video and audio and transcript also each being useful with averages ratings of between 4.1 and 4.5. A slightly smaller percentage rated the Synmark option and the print out option as being useful with an average of 3.7 and 3.3 respectively on the 5 point scale used. One student commented that since there was no experience in using these facilities they were not being used but if Synote was used regularly then it would become second nature to use all these facilities. Ninety seven percent of the students wanted all their lectures to be presented on Synote. Forty four students in two of these five classes identified in more detail how using Synote affected their learning by indicating whether Synote improved or reduced or provided no change in the aspects of their learning shown in Table 1.

Table 1. Affect on Learning

	improve	no change	reduce
learning	95%	5%	0%
attention	61%	34%	5%
motivation	50%	50%	0%
efficiency	77%	23%	0%
enjoyment	66%	32%	3%
results	69%	31%	0%
notetaking	53%	30%	18%
attendance	13%	45%	42%

A few students commented that some students' notetaking skills might not develop if all classes were recorded and presented on Synote. Some students commented that if they were ill or had serious problems they would in the past normally still have tried to come in to classes because they were worried they would get behind in their work and would be unable to catch up, whereas with Synote they could stay at home and get well secure in the knowledge that they would not miss anything as they could learn using the Synote recording. This is clearly an important issue during flu epidemics. Students stated that it was important that that ALL lectures were recorded so they didn't find that the one lecture they missed hadn't been recorded. Of these 44 students 37% identified themselves as native speakers, 26% as fluent, 28% as having good language knowledge and 9% as having little language knowledge. Non native speakers in particular commented how valuable Synote was for them as it was sometimes difficult to understand lecturers' speech and note-taking in a foreign language was very difficult for them. One commented that they could get words not understood in the transcript translated by Google. Of these 44 students 7% identified themselves as having hearing disabilities, 2% visual disabilities, 11% learning disabilities and 7% other disabilities. Two overseas students wrote "*Synote gives a very clear understanding of module ...It was very useful for me especially as I am a non English*

native speaker” and “I think Synote is a very good way to listen to lectures. If for example we miss the lecture we can actually listen to it in our own time or if we didn't really understand the lecture we can go back to it and listen to it carefully. I also like the highlighted part whenever the lecturer speaks on the text so we can't actually get lost within long texts.” Other students wrote *“Synote is very useful for students in general, I think at present all they need is to get used to something like Synote. It will then become second nature”* and *“Synote is the best system I have ever seen for assistive technology it is very useful for me to understand what the lecturer taught after class I hope all school majors could integrate this system thanks”*. Synote is used and valued by all students so non native speakers and disabled students feel more included and don't need to walk to the front of the class to ask the teacher if they can record the lecture on their personal digital recorder. Also the quality of recording from a teacher's wireless headworn microphone is significantly better than from small personal digital recorders placed at the front of the class. Students did not like retyping handwritten notes they had taken in class into Synote after the recording had been uploaded and so Synote has recently been enhanced so that synchronised notes taken live in class on mobiles or laptops using Twitter [22] [23] can be automatically uploaded into Synote. Google announced in November 2009 that they plan to use speech recognition to caption YouTube videos; a feature that has been available for nearly two years for students using Synote.

5 Conclusion

This paper has put forward the following premises and arguments: Real-time captioning can support deaf students and non native speakers; Transcripts obtained from captioning, can support disabled students who find difficulty taking notes (e.g. dyslexic, motor impaired, visually impaired, hearing impaired etc.); Manual real-time captioning using stenographers is expensive and there is a shortage of people trained to undertake this task; If captioning was seen to provide a great benefit to ALL students then the cost of captioning per student benefiting would decrease; If automatic captioning using SR could help provide captions of the required accuracy then this would help overcome the shortage of stenographers; An application such as Synote that encouraged universities to caption all audio and video recordings in order to enhance the learning and teaching benefits of multimedia recordings provides great benefit to deaf and disabled students and non native speakers. Currently available commercial SR systems (e.g. Microsoft, Nuance) do not however make a synchronised transcript output available to the user as the text and speech are only temporarily available in a proprietary synchronized format for correction purposes.

References

1. Whittaker, S., Hyland, P., Wiley, M.: Filochat handwritten notes provide access to recorded conversations. In: Proceedings of CHI 1994, pp. 271–277 (1994)
2. Wilcox, L., Schilit, B., Sawhney, N.: Dynamite: A Dynamically Organized Ink and Audio Notebook. In: Proc. of CHI 1997, pp. 186–193 (1997)

3. Chiu, P., Kapuskar, A., Reitmeief, S., Wilcox, L.: NoteLook: taking notes in meetings with digital video and ink. In: Proceedings of the seventh ACM international conference on Multimedia (Part 1), pp. 149–158 (1999)
4. Brotherton, J.A., Abowd, G.D.: Lessons Learned From eClass: Assessing Automated Capture and Access in the Classroom. *ACM Transactions on Computer-Human Interaction* 11(2) (2004)
5. Bain, K., Hines, J., Lingras, P.: Using Speech Recognition and Intelligent Search Tools to Enhance Information Accessibility. In: Proceedings of HCI International 2007. LNCS, Springer, Heidelberg (2007)
6. Wald, M.: An exploration of the potential of Automatic Speech Recognition to assist and enable receptive communication in higher education. *ALT-J, Research in Learning Technology* 14(1), 9–20 (2006)
7. Nuance, <http://www.nuance.co.uk/>
8. Leitch, D., MacMillan, T.: Liberated Learning Initiative Innovative Technology and Inclusion: Current Issues and Future Directions for Liberated Learning Research. Saint Mary's University, Nova Scotia (2003), <http://www.liberatedlearning.com> (Retrieved February 7, 2007)
9. Tegrity, <http://www.tegrity.com/>
10. ViaScribe, <http://www.liberatedlearning.com/technology/index.shtml>
11. Bain, K., Basson, S., Wald, M.: Speech recognition in university classrooms. In: Proceedings of the Fifth International ACM SIGCAPH Conference on Assistive Technologies, pp. 192–196. ACM Press, New York (2002)
12. Kheir, R., Way, T.: Inclusion of deaf students in computer science classes using real-time speech transcription. In: Proceedings of the 12th annual SIGCSE conference on Innovation and technology in computer science education (ITiCSE 2007). *ACM SIGCSE Bulletin*, vol. 39(3), pp. 261–265 (2007)
13. Bennett, S., Hewitt, J., Mellor, B., Lyon, C.: Critical Success Factors for Automatic Speech Recognition in the Classroom. In: Proceedings of HCI International (2007)
14. Wald, M.: Creating Accessible Educational Multimedia through Editing Automatic Speech Recognition Captioning in Real Time. *International Journal of Interactive Technology and Smart Education: Smarter Use of Technology in Education* 3(2), 131–142 (2006)
15. Attila, <http://www.liberatedlearning.com/news/AGMSymposium2009.html>
16. Fiscus, J., Radde, N., Garofolo, J., Le, A., Ajot, J., Laprun, C.: The Rich Transcription 2005 Spring Meeting Recognition Evaluation. National Institute Of Standards and Technology (2005)
17. Automatic Sync Technologies, <http://www.automaticsync.com/caption/>
18. Synote, <http://www.synote.org>
19. JISC, http://www.jisc.ac.uk/fundingopportunities/funding_calls/2007/07/circular0207.aspx
20. Net4Voice, <http://www.net4voice.eu>
21. Wald, M.: A Research Agenda for Transforming Pedagogy and Enhancing Inclusive Learning through Synchronised Multimedia Captioned Using Speech Recognition. In: Proceedings of ED-MEDIA 2007: World Conference on Educational Multimedia, Hypermedia & Telecommunications (2007)
22. Twitter Synote, <http://twitter.com/synote>
23. ECS news, <http://www.ecs.soton.ac.uk/about/news/2812>

Teaching English to Deaf Adults: “SignOnOne” – An Online Course for Beginners

Marlene Hilzensauer

Center for Sign Language and Deaf Communication, University of Klagenfurt,
Universitätsstraße 65-67, Klagenfurt, Austria
marlene.hilzensauer@uni-klu.ac.at

Abstract. A basic knowledge of English is indispensable nowadays. While there are many courses for hearing people, there are few aimed at deaf people who use a national sign language as their first or preferred language. This paper describes “SignOnOne”, a two-year Grundtvig project which is a follow-up to “SignOn!” (Internet English for the Deaf). “SignOnOne” teaches English for beginners with sign language as language of instruction. This allows the deaf users to access the course directly (the written language of their country is often a first foreign language). The target group are deaf adults, who can use the course in class but also explore English on their own. This online course will be accessible free-of-charge via the Internet homepage of the project without any log-in or download.

Keywords: sign language, deaf and hearing-impaired people, e-learning, accessibility, ICT, multimedia, deaf bilingualism, EFL.

1 Introduction: State of the Art and Application Idea

English is the modern “lingua franca” – without a working knowledge of English, the access to a lot of information and international opportunities is barred. This holds especially for deaf people who use their national sign language as first or preferred language. For them, English is – at best – a second language, often even a third language after the national written language of their home country.

Although there are many English courses for beginners available, most of them do not fulfil the needs of deaf learners. There are few English courses for deaf people (a rare case of how to adapt an English textbook for the use by deaf students is Hejlová [1]). Modern language teaching has shifted from the traditional grammar-translation method to focusing mainly on communication skills like speaking and listening, thereby inadvertently more or less excluding everybody who is unable to do so:

Macurová [2], for example, argues that the most important aspect of teaching language to deaf students is the fact that they cannot hear – something that is so simple that it is usually not really taken into consideration.

The deaf – defined by Macurová as “hearing impaired persons who are unable (even with hearing aids) to identify phonemes of a spoken language” [3] are barred from the acoustic channel (including teaching aids like audio CDs, MP3 files, videos with sound), therefore any information needs to be presented in a visual way. They

cannot learn any spoken language by simply acquiring it during everyday communication and interaction. This fact – in contrast to current teaching methods – lessens the importance of spoken language and emphasizes the use of written texts. Macurová stresses this conclusion of research into the teaching of the deaf, namely that “thinking and making oneself understood is more important than speech performance – and that the development of literacy of the deaf is ultimately more important than the development of spoken language.” [4]. In our opinion, this means that English learning should at least start with written language, taught with sign language as language of instruction. By using sign language videos, deaf learners can access the information directly. There is no danger of misunderstanding something because of translations from one written language into another one (as often even the written language of their home country is a second language for the deaf). Because there is no language difficulty, it gives them the option to work with the course on their own, without the presence of a teacher. Furthermore, the use of sign language respects their preferences and their culture.

2 The “SignOn!” Method: Written Text and Sign Language

The first step towards sign language based learning for the deaf was the “SignOn!” project (Socrates Lingua Action 2, 2004-2007), which was aimed at deaf adults with basic English skills. The outcome of “SignOn!” is a multimedia course for Internet and international English. This online course is accessible via the project homepage (www.sign-on.eu). It comprises ten lesson texts dealing with Internet and deaf topics. These “Main Texts” give an overview of the topics and introduce the most important words and phrases. There is no specific order in which to work through the texts; the navigation within the course is completely up to the users themselves.

Next to the written text, sign language translations are available in video form throughout the course, both in the national sign languages of the seven partner countries (Austria, Finland, Iceland, the Netherlands, Norway, Spain and the United Kingdom) and partially also in International Sign. When entering the course, the users have to choose a sign language first; they can switch to another one at any time. They can select between different “tabs” for different kinds of translations: sign language videos are available for the complete texts, single sentences and difficult words and phrases (selected by the deaf team members of the project partners), as well as for the introduction to the course.

The users can also test their skills with interactive exercises or look up brief signed grammar explanations (because the target group are intermediate/advanced learners, these explanations are not all-encompassing, but rather summaries of the most important points). A more detailed description of the program can be found in Hilzensauer and Skant [5].

2.1 The Results of “SignOn!”

The feedback to “SignOn!” showed good acceptance by the users as well as an unexpected additional bonus: as the users can choose between different sign language translations, the program may also be used to learn signs or phrases from other sign

languages. One fact that emerged during the dissemination was the urgent need for a course for beginners. Both potential deaf users and teachers of the deaf argued that “SignOn!” – as an intermediate course – was too difficult for complete beginners and that they would like to have a basic English course using the same method.

3 The “SignOn!” Method for Beginners: “SignOnOne”

The results of the feedback to “SignOn!” led to a follow-up project, “SignOnOne” (2008–2010), which is an English course for deaf adults without any knowledge of English at all. There were some changes in the partner consortium: Finland, the Netherlands and the United Kingdom had to leave due to other commitments, while the Czech Republic and Hungary were new additions. Austria is the coordinator of both projects, while Norway is responsible for the programming. The project has been funded with the support of the Grundtvig programme of the European Union and the Austrian Federal Ministry for Education, Arts and Culture.

3.1 Basic Design of “SignOnOne”

The basic structure of the course with ten lessons centering on the “Main Texts” was kept. The navigation is still free, although it is strongly recommended to work through the lessons in the order in which they are presented as the lessons build on each other.

The screenshot displays the 'Main Text' of Lesson 3 in the SignOnOne course. The interface is organized as follows:

- Header:** 'Theme: Seasons' and a navigation bar with buttons for '1', '2', '3' (selected), '4', '5', '6', '7', '8', '9', '10', 'Word list', and 'Grammar reference'.
- Sub-navigation:** A row of buttons for 'Animation', 'Sentence' (selected), 'Word', 'Phrase', 'Grammar', and 'Exercise'.
- Main Text Area:**
 - Spring:** A paragraph describing the first season of the year, listing months (March, April, May), colors of flowers (red, yellow, white, blue), and the state of trees and grass.
 - Weather:** A paragraph stating that the weather is warm but rainy, with sunny periods and black clouds.
 - Easter:** A paragraph describing the Easter story, mentioning chocolate eggs, a brown bunny, and children wishing 'Happy Easter!'
- Video Player:** A video window showing a man in a black shirt signing. Below the video is a progress bar and a row of language selection icons (Austria, Czech Republic, Hungary, Norway, Denmark, and a smiley face).
- Footer:** 'levels nav' and 'Page 1 of 4'.

Fig. 1. Demo version: “Main Text” of Lesson 3

The contents of the ten lessons were developed in close cooperation with the deaf team members of each partner country. The lesson texts do not contain an ongoing story, but the lessons are loosely connected through the same protagonists: a family with some deaf family members as well as their friends and relatives. The topics – again

chosen by the deaf collaborators – deal with basic situations like introducing yourself to someone, weather and seasons, free time activities, etc. They teach the basic vocabulary necessary to be able to function in everyday life as well as special words and phrases connected to the topics (approximately 1000 words for the whole course).

The design with six “tabs” – retained from “SignOn!” with a few small changes – enables the users to identify consecutively smaller units within the texts and to form and check their own hypotheses regarding their meaning.

The first tab is an innovation; instead of the translation of the whole text it gives a visual representation of the content of each sentence – if the users click on any sentence in a lesson text, they will get an animation which allows them to identify the most important cognitive concepts within this sentence. These animations are independent from any specific language, whether written or signed.

The screenshot shows the SignOnOne web application interface. At the top, there is a navigation bar with tabs for "Animation", "Sentence", "Word", "Phrase", "Grammar", and "Exercise". The "Animation" tab is selected. The main content area displays text about Spring, including a paragraph about the weather and a paragraph about Easter. A video player on the right shows a hand-drawn diagram of a semi-circle divided into three segments labeled "March", "April", and "May", with the word "spring" written across the segments. The interface also features a logo for SignOnOne and a footer with the European Union flag and the text "Education and Culture DG Lifelong Learning Programme".

Fig. 2. Example for an animation (the three months in spring)

The second tab contains a signed translation of each sentence, the third and fourth the translation of each word or phrase respectively (alternatively, if a direct translation is impossible, a signed explanation). The users may select the sign language by clicking on the flag of the respective country. There is a translation for every word in the text, even for those which make up a phrase, so that the learners can compare the meaning of single words with the different meaning given by the phrase. All the words and phrases can also be found in an alphabetical word list. The grammar terms used in the grammar explanations will also be translated and/or explained in sign language. This will help the learners to understand these terms and enable them to work with grammar books.

Via the “grammar” tab, the user has access to grammar tables and signed grammar explanations. The grammar will be presented as visually as possible, e.g. using a flip chart with examples in the videos.

The screenshot shows the 'Grammar' tab of the SignOnOne course. The theme is 'Family'. The navigation menu includes 'Animation', 'Sentence', 'Word', 'Phrase', 'Grammar', and 'Exercise'. The 'Grammar' tab is active, displaying a table for the 'To be' verb in the negative form. To the right, a video shows a sign language instructor pointing to a flip chart with the following text:

I am not nice.
You are not nice.
You aren't.
You're not.

The table below is a reproduction of the one shown in the screenshot:

To be	Negative	Contractions	
Singular			
I am	not	I am not	I'm not
you are		you aren't	you're not
he is		he isn't	he's not
she is		she isn't	she's not
it is		it isn't	it's not

The interface also features a 'sign on one' logo, a video player with a progress bar, and a row of flags representing different languages. At the bottom, there is a logo for the European Union and the Lifelong Learning Programme.

Fig. 3. Example for a signed grammar explanation, using tables and a flip chart

A separate “grammar reference” gives an overview of all the grammar phenomena the users will encounter throughout the course. Although less detailed than the grammar explanations within the individual lessons, the users will find here tables and examples in a logically structured menu which allows for quick and easy orientation.

The last tab, “Exercises”, contains video exercises, e.g. matching a sign with the respective written English word, cloze tests, and arranging the words of a sentence or the letters of a word in the correct order (with the help of the signed translation). The exercises contain sign language videos and/or pictures to make them visually more interesting. The instructions for the exercises are available in both written and signed form: by clicking on the question mark behind the English instructions the user opens a new window for the sign language video. So if the users want to repeat an exercise or are already familiar with this exercise type, they are not forced to watch the same video over and over again. There are no explicit grades; the users are simply told whether their answers were correct or not by using visual symbols like a green checkmark, a red X or an emoticon. The number of correct answers is not stored; neither is the number of mistakes.

A special feature of “SignOnOne” is what we nicknamed the “Talking Head”: while “SignOn!” did not use any audio, the new course includes videos of the face of a native British speaker pronouncing all the sentences, words and phrases. This was included because the deaf are interested in the correct pronunciation – as the English pronunciation often differs widely from the written form and can be confusing, they would have trouble understanding other people or talking in English themselves. Deaf

users can use these videos to lip-read English, while hard-of-hearing users might even be able to work with the actual sound files. At the moment, this feature is activated by clicking on the button with a Smiley on it; the Smiley will be changed to a picture of lips or the lower half of a face (the exact symbol is still being debated by the deaf collaborators).

The screenshot shows a web interface for 'SignOnOne Exercises'. At the top left is the 'signonone' logo. To its right is the title 'Exercises' and the text 'School, work, education / Multiple Choice 2'. On the far right are three links: 'Close exercises', 'Back to exercises menu', and 'Click on the correct word to fill the gap'. Below the header is a video player showing a woman signing. Below the video is the text 'Steven ___ Peter his computer.' To the right of the video are three multiple-choice options: 'talked', 'showed', and 'entered'. At the bottom right of the interface is a page indicator 'Page 2 of 5'.

Fig. 4. Example for an exercise with a sign language video (cloze test)

“SignOnOne” will also include a signed help which is divided into two parts: the first one is an introduction which welcomes the users to “SignOnOne” and gives a general explanation of the contents and how to work with the course. The second part explains the different features in detail; the respective sign language videos can also be called up separately if the users want to look up some specific information. This help mode will probably be accessed via a help button; the exact way of doing this is still being discussed by the partners.

3.2 Technical Background

Originally, we had planned to re-use the HTML program from “SignOn!” for “Sign-OnOne”, but – following a proposal by the Norwegian partners – we decided on new programming in Flash ActionScript instead, which should be working with all common web browsers.

An important issue was that we wanted to have an option to extend the program to include additional texts or sign languages and the programming had to allow for this. Therefore all the texts and the linking information are stored in external text files and XML files. The linking has to be the same for every sign language and all the sign language video files for each country are contained in a single folder. This makes it

quite easy to reuse the interface with different texts and video files. The grammar and the exercises are programmed in Flash; for reusing these, some knowledge of Flash is necessary.

3.3 The Results of “SignOnOne” So Far

After the first half of the project duration, the first five lessons of “SignOnOne” are already finished and implemented in a demo website. This shows that the methodological concept can be realised in practice. The partners are now working on the second half of the contents (the contents are mostly finished and the filming of the sign language videos has already begun). The first dissemination activities during which the demo website was presented have shown great interest in the finished product.

4 The Impact of “SignOn!” and “SignOnOne”

The unique approach of both “SignOn” courses is that they try to give the learners the freedom to experiment with language on their own. Although we recommend using “SignOnOne” as a course for beginners together with a teacher, the deaf students can try to understand the texts, learn new words and phrases or do the exercises on their own. Besides, we believe like Schumaker [6] “that students become more interested in their own development in English when they assume the responsibility for their own learning.” People have different learning needs and learning strategies, therefore the course allows them to proceed on their own as much as possible. They can spend as much time as they want to on each part of the program. Each step may be repeated as often as they like without having to fear that they will get bad marks for making mistakes or advancing too slowly.

Furthermore, “SignOnOne” will provide teachers with the opportunity to use an English course for deaf beginners which was specifically developed for this target group. The participation of deaf team members from each partner country guarantees a permanent deaf perspective; these are courses designed by deaf people for other deaf people and not something conceived by hearing people according to their preferences.

5 Conclusion and Planned Activities

Both “SignOn” courses can be accessed via their Internet address. They are free of charge and very easy to handle – the users do not need to download and install anything and they do not need any computer skills other than clicking on buttons or using the menu of the video player. They are also anonymous: there is no log-in necessary and the program does not keep track of what the users are doing. We decided to avoid the usual design with entrance exams, tests and grades because this is too reminiscent of school and might deter the deaf users.

Due to the Flash programming, it will be possible to add new contents or new sign languages to “SignOnOne” without advanced programming skills. If other countries should be interested in this course (there have already been first requests), they only have to provide the sign language videos and to adapt the grammar explanations to

their own written language. It will also allow us and any other interested parties to develop new lessons at any time. While “SignOnOne” teaches only the basics and “SignOn!” the special terminology needed for the Internet and international communication, the final result should be a complete English course for the deaf – one that enables them to learn English through their own national sign language so that they can understand written (and maybe even spoken) English and write it themselves.

Although this is a beginning, the scope of “SignOn!” and “SignOnOne” is still too narrow – in order to provide deaf people with the same chances as hearing people, they need access to the same amount of information as the latter. This means that complete courses for hearing people would have to be adapted with the “SignOn” method to make them barrier-free.

References

1. Hejlová, H.: *Headway Textbooks for Deaf Students*. Federation of Parents and Friends of Hearing Impaired, Prague (2003)
2. Macurová, A.: *An Introduction to Principles of Language Learning*. In: Janáková, D. (ed.) *Proceedings 2004. Teaching English to Deaf and Hard-of-Hearing Students at Secondary and Tertiary Levels of Education in the Czech Republic*, pp. 28–41, p. 28. Eurolex Bohemia, Prague (2005)
3. *Ibid.*, p. 31
4. *Ibid.*, p. 29
5. Hilzensauer, M., Skant, A.: *SignOn! – English for Deaf Sign Language Users on the Internet*. In: Kellett Bidoli, C.J., Ochse, E. (eds.) *English in International Deaf Communication. Linguistic Insights: Studies in Language and Communication*, vol. 72, pp. 155–177. Peter Lang, Bern (2008)
6. Schumaker, P.F.: *Fishing for English: Teaching Strategic Approaches to Learning English*. In: Janáková, D. (ed.) *Proceedings 2000. Teaching English to Deaf and Hard-of-hearing Students at Secondary and Tertiary Levels of Education in the Czech Republic (2nd printing)*, pp. 144–147, p. 146. Eurolex Bohemia, Prague (2005)

A Web Based Platform for Sign Language Corpus Creation

Davide Barberis, Nicola Garazzino, Elio Piccolo, Paolo Prinetto,
and Gabriele Tiotto

Politecnico di Torino, Dipartimento di Automatica e Informatica,
Corso Duca degli Abruzzi 24,
10128 Torino, Italy
{davide.barberis,nicola.garazzino,elio.piccolo,
paolo.prinetto,gabriele.tiotto}@polito.it

Abstract. This paper presents the design and implementation issues of a tool for the annotation of sign language based on speech recognition. It is at his first version of stable development and it was designed within the Automatic Translation into Sign Languages (ATLAS) project. The tool can be used to gather data from the language for corpus creation and stores information in a Multilanguage Multimedia Archive. The objective is to create a corpus of Italian Sign Language for supporting an automatic translation system that visualizes signs by means of a virtual character. The tool allows the user to search for media content related to signs and stores metadata about the translation and annotation processes.

Keywords: Accessibility, Sign Language, Annotation, Language Formal Representation, Annotation Editor.

1 Introduction

Deaf people represent 0.1% of the Italian population (about 50,000), however, 10% of the Italian population suffers from relevant hearing impairment [1]. This percentage is increasing of 4%-6% every year. Severe deafness represents about 25%-30% of the cause of invalidity (500,000 deaf in Italy).

The integration difficulties of persons who were born deaf or acquired deafness in the first years of life are higher, because they could not acquire the knowledge of the spoken language. It is reported that just a small percentage of the deaf persons has an acceptable knowledge of the written Italian language. Sign Languages allow deaf children to acquire a full cognitive development within their community composed of hearing people and deaf persons. Such cognitive development represents the basic prerequisite to a full access to the education, the culture, and the inclusion in working and social environments.

The depicted scenario makes evident the relevance of the availability of a low cost system to cover, in a sustainable way, the increasing request of a wide range of services including Italian Sign Language contents.

Many projects targets sign language visualization and more in general accessibility to deaf people, and an increase of interest in this research field is reported in the last years. DePaul University ASL synthesizer allows the visualization of signs in ASL [2], Signspeak [3] targets sign language recognition and translation, the DICTA-Sign project targets sign language recognition and synthesis. are examples of such projects. As demonstrated by the results of some international projects [4][5], virtual character based solutions are generally accepted by the deaf community and while some works revealed that this technology could give a satisfying level of understandability [6], additional work has to be done to improve the translation quality, in order to provide a sign language translation resorting to its grammar. Younger people are very interested in this technology, since it offers a new possibility of being included in the global community: a natural evolution of a process started by computers, the internet, and mobile phones. To implement a virtual character that signs as similar as possible to a real human, it is necessary to study the language by a technical and linguistic point of view. These studies will be based on a corpus of sentences in Italian Language and Italian Sign Language (LIS) created by movies of signing humans. It is well known that data gathering through annotation of these movies is a complex and hard task. For this reason the use of tools to provide assistance to the operator can be useful and improves quality in annotation. Several tools for video annotation are present and freely downloadable from the internet, such as ELAN [7] or ANVIL [8]. They are mature and useful in several research areas. It is possible to annotate Sign Language videos with them but to our knowledge no tool targeting Sign Language annotation is available. In this paper we present a tool, called ATLAS Advanced LIS Editor for Annotation (A^2LEA). It has been designed for the annotation task and built around Sign Language annotation needs. It implements features to aid the operator in the annotation task. In chapter 2 we describe the analysis of LIS that we performed and that is on the basis of the tool. In chapter 3 a description of the tool is provided. In section 4 a description of the annotation methodology is given with an example. The paper concludes with future developments.

2 About Data: The Analysis of Language

The availability of a corpus is a necessary requisite for any research study that has the objective of gathering linguistic resources. Various Corpora exists for many spoken languages and some work has been done for sign languages. To our knowledge no corpus of LIS actually exists and sign language, even if it is a real language with its own grammar and morphology, has a not well-established reference grammar. The creation of a corpus is therefore needed for linguistic research purposes and has been started within the Automatic TransLation into Sign Languages project (ATLAS) [9]. The ATLAS project targets the development of a platform for the automatic translation from Italian to LIS in the weather forecasts domain. This is a relatively new domain for sign language since it suffers from a considerable lack of lexicon. Texts are prepared resorting to a real weather forecasts Italian texts corpus and terms that are not present in sign language are detected. Moreover a lexicon of such terms has been created that includes video recordings of the single signs. Video translations have been provided including these signs in utterances. From the analysis

of these videos, was possible to identify peculiarities of LIS in standard signs and in weather forecasts new signs. The analysis of the video has been performed from the linguistic point of view and from the animation point of view. By this analysis a considerable number of grammar phenomena was detected. These phenomena allowed us to code sign elements and to build a formal language that conveys both linguistic and visualization information. This is called ATLAS Extended Written Italian Sign Language (AEWLIS). A description of the ISL peculiarities and of the corpus construction methodology is out of the scope of this paper but it is important to point out that A²LEA is built on the AEWLIS definition.

3 The System Architecture

In figure 1 the system architecture is shown. A²LEA is a web application, so that is possible to annotate by simply connecting to a website. This allows what we define a “distributed annotation” as an evolution of the collaborative annotation that uses peer to peer technology [10]. Many users can take effort of the same media content and produce an annotation. The system stores user information and ranks it on the base of his experience as annotation operators and on other criteria such as: experience as a linguist, LIS knowledge level, computer skills, etc...

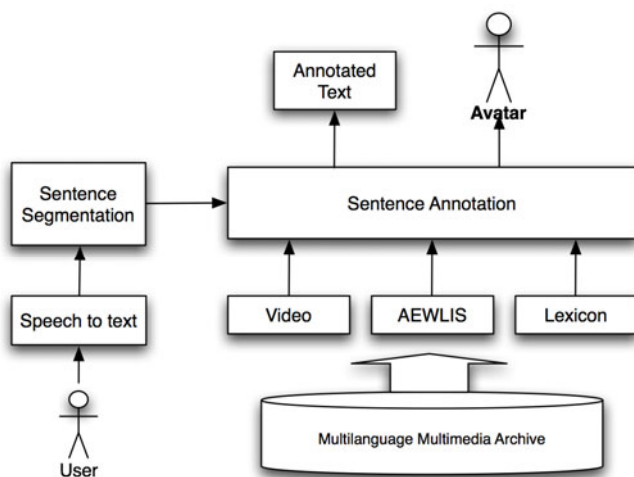


Fig. 1. Annotation Tool Architecture

A²LEA takes effort of a speech to text recognizer. If the user is a LIS expert, he can load the video content and just “speak” the lemmas as he sees them in the video. The output is a sort of written LIS that can be loaded by the sentence segmentation module. This produces the single sentence to be loaded by the Sentence Annotation Module (SAM). This is the core of A²LEA, it loads the segmented sentences and provides a web interface for the operator to perform annotation. A screenshot of the

Sentence Annotation Module (SAM) is provided in figure 2. SAM loads the video information and allows the user to performs the following operations:

- Sentence modification (lemmas additions and deletions, order changes),
- Lemma tagging (association with attributes proper of the AEWLIS specification and format),
- Linking to lexicon resources (linking each lemmas to a sign in the LIS dictionary).

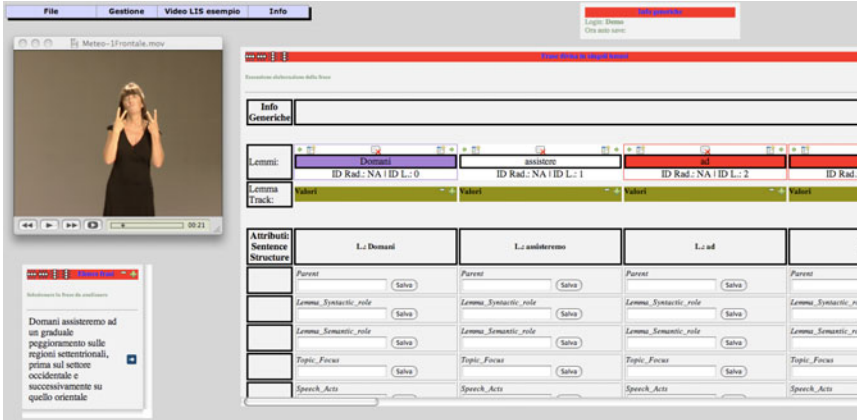


Fig. 2. Sentence Annotation example

Each of these operation implies a continuous interfacing with the Multilanguage Multimedia Archive (MMA). The annotation is performed by modifying the sentence and by tagging the lemmas. As a consequence the annotation is at the level of sentence defined as a concatenation of utterances, and at the level of lemmas. The third operation takes effort of the first consistent online LIS dictionary that has been built within the ATLAS project. It is based on the Radutzky Dictionary [11], that has been digitalized through a custom user interface and stored in the MMA. For each coded sign it stores a video and a short example of the sign usage. This information is available online through A²LEA. As we pointed out in chapter 2 the AEWLIS formalism has a strong visualization component that allows the user to describe the sign that he can see in the video. Therefore it is possible to specify the position in which each sign is performed even with respect of the other signs (called entities in the signer's space). The A²LEA will help the user in this complex task by providing an image of the signer (referred as Avatar in picture 1) with coordinate points in the signer's space. This feature will be embedded in the Sentence Annotation Module.

4 Annotation Methodology

A²LEA takes as input a text file, that can be produced by the speech recognizer or eventually loaded by the user. The tool provides also manual text insertion, in this way a LIS expert can write the lemmas that he sees in the video directly in a text box.

The input text is analyzed by means of a morphological analyzer [12] and the relevant lemmas are properly pointed out. In the following we have an example in which the text in input is “La mela è sopra il tavolo” (the apple is on the table). The correspondent analysis is performed as shown in the table below.

Table 1. Annotation example of the italian sentence for “the apple is on the table”

Italian	La mela è sopra il tavolo					
Position	1	2	3	4	5	6
Step 0	LA	MELA	ESSERE	SOPRA	IL	TAVOLO
Candidate for deletion	Yes	No	No	No	Yes	No
Step 1		MELA		SOPRA		TAVOLO
Step 2		MELA		TAVOLO		SOPRA
Step 3		Topic Agent(3)		Focus Location(3)		

The morphological analyzer extracts useful information (i.e. number and syntactic role) on the lemmas, that have impact on the signing act. By means of the syntactic role some lemmas are identified as candidates for the deletion as they will not constitute signing elements on the basis of the ISL grammar. Notice that the verbs are not candidates but in this case “essere (to be)” will be manually deleted in step 1. In this step all the candidates are automatically deleted. In step 2 the lemmas are manually shifted by the user according to the signing act that is showed on the video. In step 3 lemmas attribute are added. The attribute are the same specified in the AEWLIS formalism. In table 1 two of them are reported:

- The Topicalization, which is necessary for the identification of the main element of the sentence. The corpus creation phase revealed that the order of the elements in the sentence changes in sign language and, even if the standard form is SUBJECT-OBJECT-VERB, this phenomenon can be found in ASL as well. Hence in some cases we have the same sentence signed in different ways by different deaf persons. It became clear that is necessary to identify what is the topic and the focus as it influences the signing act, in particular topicalization is realized with use of non manual component in correspondence of the topic.
- The Thematic Role, which is useful to identify semantic relations between the element of the sentence. This attribute is important to build the scene, that is locating the entities in the scene, where entity is an abstract element that becomes a concrete sign within the scene (including classifier). In some cases entities are placeholders to manage signs movements in the space.

The signed sentence could have more entities than the number of lemmas present in the annotated sentence after step 3. In this case it is possible to add those elements according to the LIS grammar. Hence, the output AEWLIS sentence will result as reported in table 2.

Table 2. Annotation example with additional lemmas

Italian	La mela è sopra il tavolo			
Position	1	2	3	4
Step 2	MELA	TAVOLO	CLASS(plane)	HANDSHAPE (C)
Step 3	Topic Agent(3)	Focus Location(3)		
				Place_top(3)

This is a more accurate annotation. In LIS the “SOPRA” lemma which means “is on”, is signed by first placing a classifier representing plane things in the neutral space with the left hand, then by performing the “C” handshape (that is the handshape used to perform the “MELA” –apple sign) with the right hand and by placing it on the classifier. Thus the iconicity is resolved by physically placing the apple on the table. This can be annotated with a lemma of type “classifier” and a lemma of type “handshape”. In the step 3 the location attribute is added. This specifies that the lemma in position 4 is spatially placed on the top of the lemma in position 3.

The annotation performed in the example above is just a partial annotation with respect of the amount of attributes defined in the AEWLIS formalism. It includes, among others, facial expressions, hands specification (left, right, both), residual (a frequent phenomenon in which the signer keeps the hands stopped at the end of one sign while signing with the other hand), movements of the body, etc... The AEWLIS formalism is still under refinement since new LIS phenomena and peculiarities about its grammar are constantly found.

5 Conclusion and Future Developments

In this paper we presented a new tool for the annotation of videos completely sign language oriented. The advantages of using this tool are several. It speeds up the annotation by acquiring the lemmas through a speech to text module. It automatically parses this input and provides an interface in which the user can easily insert the attributes of each lemma from a linguistic and visualization point of view. We also introduced the paradigm of distributed annotation brought up by the web based architecture. Future developments include the customization and integration of the speech to text recognizer, the development and integration of the sign position and trajectory input module and the performance trials. An attempt will be made to train the recognizer in order to identify the other special kind of lemmas (i.e. handshape or classifier, see section 4), by means of keywords. This feature can be extended to automatically add non-manual features like facial expressions These can be conducted by different users with different backgrounds and annotation timings and accuracy can be measured. Comparisons with different annotation tools can be performed to detect problems and possible improvements.

References

1. Medical Survey,
<http://www.numedionline.it/numedi/document/comairs.html>
2. Schnepf, J.: Towards Computerized Synthesis of Nonmanual Signals in American Sign Language. In: DePaul CTI Research Symposium (CTIRS) & Midwest Software Engineering Conference (MSEC), Chicago, Illinois (2006)
3. <http://www.signspeak.eu/en/index.html>
4. Efthimiou, E., et al.: Sign Language Recognition, Generation, and Modeling: A Research Effort with Applications in Deaf Communication. In: Proceedings of UAHCI 2009/HCI 2009. LNCS, vol. 5614, pp. 21–30. Springer, Heidelberg (2009)
5. Signing Science Dictionary homepage, <http://signsci.terc.edu/>
6. Igi, S., et al.: Sign language synthesis for mobile environments. In: Proc. of the 11th International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision (February 2003)
7. Brugman, H., Russel, A.: Annotating multi-media / multi-modal resources with ELAN. In: 4th International Conference on Language Resources and Evaluation, pp. 2065–2068. European Language Resources Association, Paris (2004)
8. Kipp, M.: Anvil - A Generic Annotation Tool for Multimodal Dialogue. In: Proceedings of the 7th European Conference on Speech Communication and Technology (Eurospeech), pp. 1367–1370 (2001)
9. <http://www.atlas.polito.it>
10. Brugman, H., Crasborn, O.A., Russel, A.: Collaborative annotation of sign language data with peer-to-peer technology. In: 4th International Conference on Language Resources and Evaluation, pp. 213–216. European Language Resources Association, Paris (2004)
11. Mason Perkins Deafness Fund. Dizionario Bilingue Elementare della Lingua dei Segni Italiana LIS. (Ed.) Kappa. Rome, Italy (2001) (in Italian)
12. Pianta, E., Girardi, C., Zanoli, R.: The TextPro tool suite. In: Proceedings of LREC, 6th edition of the Language Resources and Evaluation Conference, Marrakech, Morocco (2008)

Context Analysis of Universal Communication through Local Sign Languages Applying Multivariate Analysis

Naotsune Hosono¹, Hiromitsu Inoue², and Yuji Nagashima³

¹Oki Consulting Solutions Co., Ltd.,
4-11-15 Shibaura, Minato-ku, Tokyo 108-8551 Japan

²Chiba Prefectural University of Health Sciences

³Kogakuin University
hosono903@oki.com

Abstract. This paper discusses universal communication with ICONs or pictograms in the field of assistive technology (AT) with human centred design (HCD) and context analysis by Persona model. Typical two personas are created as a deaf person in an emergency and a travelling woman from Hongkong. Then scenarios like diary are written and about 40 words are selected as minimum communication words in the dialogue. Several local sign languages of related selected words are referred in order to investigate universal signs. For this purpose a sensory evaluation method with multivariate analysis is applied. The outcome is plotted on one plane with relationships of subjects and samples of several local sign languages. Through proposed method by sensory evaluation, the relationship between fundamental words and local sign languages are initially explained.

Keywords: assistive technology (AT), human centred design (HCD), context of use, Persona model, sign language.

1 Introduction

This research was started in order to investigate the context of universal communication through local sign languages applying Multivariate Analysis (MVA).

One of the authors visited Rio de Janeiro in 2005 with a deaf Japanese colleague to attend the International Conference on Universal Design organized by Adaptive Environments, now renamed as Institute for Human Centred Design. Although Japanese sign language is quite different from the Brazilian one, deaf attendees were instantly creating signs and quickly communicating with each other. After observing this situation, the original research idea was conceived to find some common signs that will form a fundamental communication method with pictograms and icons for not only the deaf but also foreign travelers as well.

This research is supported and funded by SCOPE (Strategic Information and Communications R&D Promotion Program) project organized by Ministry of Internal Affairs and Communications (MIC) of Japan.

2 Background

Recent discussions in the field of ergonomics are focused on Human Centred Design (HCD) especially with regard to usability and accessibility of products [1]. The concept became an international standard as ISO 13407: 1998 Human-centred design processes for interactive systems, and now is under review as to be ISO 9241 Part 210. It has been applied to the development processes by many manufacturers bringing with it the innovative concept of making production plans and designs not from product resources (seeds) but from user requirements (needs).

HCD is based on the context of use and now standardized as ISO 9241 Part 11: 1998, Ergonomic requirements for office work with visual display terminals (VDTs) - Guidance on usability. The context of use is organized by four factors as user, product, task and environment in use. The research scope covers not only linguistic studies of sign language but also HCD with context of use.

3 Determining the Context

Based on the concept described above, two context situations have initially been chosen. Alan Cooper proposed the Persona Model [2] related to HCD where several situation representing Personas are imaginably created in order to simulate and find how they will behave under a certain context. This method is highly accepted by modern companies in creating new product plans and has been applied to service science as well.

Applying this method, the first phase is to create two Personas. The first one is a deaf person in a situation where he suffers a sudden illness while commuting in the morning, and is carried to the hospital by ambulance. The second one is a woman who lives in Hong Kong and has to visit Tokyo on business. Afterwards, she will enjoy some free time in the locales of Japan.

The second phase is to create diary like scenarios to describe their daily life. These scenarios mainly pay attention to the dialogues between the Persona and those people surrounding them. The concept of the dialogue is referred to ISO 9241 Part 10: 1996, Ergonomic requirements for office work with visual display terminals (VDTs) -- Dialogue principles. The first scenario of the deaf person in an emergency consists of about 600 words (equivalent to 3000 Japanese characters) and second with the traveling woman about 1700 words (equivalent to 8500 characters). Two scenarios were created from discussions with three colleagues utilizing the Brain Storming Method.

The third phase is to select words that are fundamentally essential to the dialogues of the scenarios. 37 words were selected and categorized as follows:

Ten greeting words that trigger the initial dialogues:

thank you, hello, good-bye, indeed, yes, no, I am deaf, do not understand, sorry, please do.

The next step of the dialogues start commonly with seven interrogative pronoun words:

where ?, how much ?, how many ?, when ?, what ?, which one ?, Who ?.

12 associated reply words for the inquired interrogative pronouns are:

toilet, country, numeric (0, 1, 2, 3, 5, 10), yesterday, today, tomorrow, name.

Three essential adjectives are found in the scenario:

painful, different, expensive.

Five essential verbs are:

want to, go, come, buy, reserve.

Looking at the dialogues in the scenarios under the selected context, the hardest process is initiating the dialogue. In modern times, people are worried about security. They are extremely cautious when approached by an unfamiliar person. Staffs at hotels or railway stations are not upset since they are used to receiving inquires from strangers, however dialogues are harder to start on the open streets. Several interjections are included to assist the initiation of dialogues. This research is focused upon dialogues with several participants and referring to observations from the viewpoints of the provider and the receiver.

From the aspect of dialogues, the seven fundamentals interrogative pronoun words are as “where?, how much?, how many?, when?, what?, which one?, who?”. The word “why?” is excluded from the first experiment since it results in a broad array of answers. Reviewing the dialogues, common answers to the interrogative pronouns tend to be replied by just a collection of single nouns.

Noun such as an “apple” is easy to engage in a conversation by simply showing its picture or photo. There is a well known booklet for this purpose titled “point it” [3] which is sold at not only bookshops but also tourist offices for the benefit of foreign travelers. On the other hand, complicated conversations and proper nouns are better expressed as written messages. Abstract nouns such as “beauty” or “simplicity” are often related to adjectives, and their expressions are an issue. In addition, numeric variations are quite often used in normal conversations. Numbers such as Zero, One, Two, Five, and Ten are selected at the first experiment and longer numbers will be written in messages.

Adjectives, similar to abstract nouns, are also difficult to express and “painful”, “different” and “high” are included in the first experiment.

4 Experiments

The research is initially focused upon pictograms to make dialogues since the fundamentals of sign language are “shape”, “position” and “direction”. This research references to a collection of figures consists of seven local sign languages [4] whose author is a deaf architect, gave overwhelming support to the research by supplying and permitting reference to the database. The seven local sign languages are of American, British, Chinese, French, Korean, Japanese, and Spanish.

In the experiment, subjects are first shown an expression with the collection of figures consists of seven local sign languages. Then subjects are asked to imagine what message is explained from those signs. After subjects are informed of the sign meaning, they are requested to vote with 19 tokens which of the seven different local sign language expressions best coincides with their image. They are permitted to put up to seven tokens for one sign language expression (Fig. 1). This sensory evaluation method can easily make relative comparisons between the seven expressions of local sign languages and is more applicable than the ordering method or pair comparison method [5].

Then the correspondence analysis of Multivariate Analysis (MVA) by statistic Software, Statistical Package for Social Science (SPSS) [6] is applied. The outcome is plotted with two axes of first and second eigenvalues. The advantage of the analysis is that the similar features of subjects and samples, in this case local sign languages, are to be plotted closely on a plane. The first experiment subjects are 13 people in their 20's including nine science course students, four humanity course students. Some have experience living overseas.

After voting by the tokens, all the subjects are asked of their confidence level with Semantic Differential (SD) method.



Fig. 1. A view of sensory evaluation method by subjects

5 Conclusion and Discussions

This paper discusses a method to extract the summarized expression of several local sign languages in order to draw pictograms or icons by applying the sensory evaluation with MVA. This method is valid in practice since Japanese and Korean sign languages are similar by historical background, and in fact they are plotted close to each other. In the characteristics of correspondence analysis, the subjects who have general and standard ideas are positioned in the centre, whereas those who have extreme or specialized ideas are positioned away from the centre. The center crossing point of the first and second eigenvalues is gravity point or average point.

The result plots of “when?” are all positioned relatively close forming cluster pattern (Fig. 2). This is because it must be easy to initiate the dialog with the expression of a wristwatch. The plots of “hello” are positioned into two cluster patterns with American, British and Spanish in one, and Chinese, Korean and Japanese in the other. The plots of “painful” are positioned into three cluster patterns with American and Spanish in one, British and French in a second, and Chinese, Korean and Japanese in a third. Subjects with overseas experience tend to be positioned relatively in the middle.

Through the proposed method, the relationship between selected words and local sign languages are initially explained by sensory evaluation of the subjects. They chose several favorable sign language expressions. The next step is for designers to draw pictograms or icons referring to those outcomes. Then under the cycle of HCD, for future designers summarized by the sensory evaluation mentioned above will be performed adding the newly designed pictograms. When the summarized one to be designed by the outcome is positioned in the middle of the clusters then it is representative of the clusters. The acquisition of user experience is to include it as a design guideline for context of emergency and traveling situations.

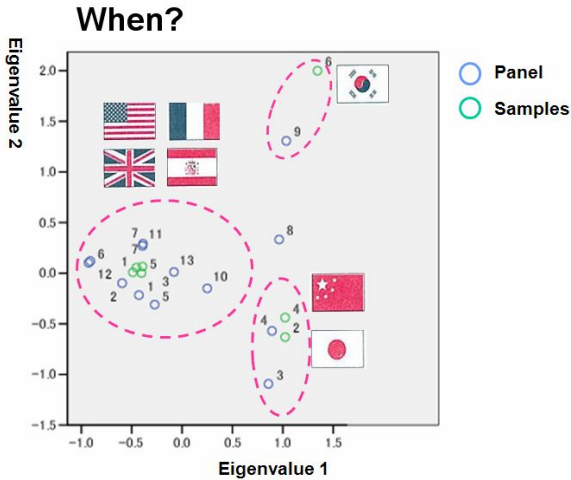


Fig. 2. The result plots by MVA of “When?”

Acknowledgments. This research is funded by SCOPE project of MIC of Japan. A collection of local sign language database is supplied and permitted for use by Mr. Akatsuka, M. of Architectural Association of Japanese Deaf (AAJD). Special thanks to Prof. Dr. Okamoto, A. of Tsukuba University of Technology (NTUT) and Dr. Nakazono, K. of NTT Network Innovation Laboratory who gave valuable comments for the project.

References

1. Miki, H., Hosono, N.: Universal Design with Information Technology (Japanese version), Maruzen, Tokyo (2005)
2. Cooper, A.: About Face 3. Wiley, Indianapolis (2007)
3. Graf, D.: point it, Graf editions, Muenchen (1992)
4. Akatsuka, M.: Seven sign languages for tourists: Useful words and expressions. In: Chinese-Japanese-American Working Group, Osaka (2005)
5. Hosono, N., Inoue, H., Tomita, Y.: Sensory analysis method applied to develop initial machine specification. Measurement 32, 7–13 (2002)
6. Field, A.: Discovering Statistics Using SPSS, 3rd edn. Sage Publications, Thousand Oaks (2009)

Toward Automatic Sign Language Recognition from Web3D Based Scenes

Kabil Jaballah and Mohamed Jemni

High School Of Science and Techniques of Tunisia,
5 avenue Taha Hussein. BP 56, Bab Mnara, Tunisia
UTIC Research Laboratory

Kabil.jaballah@utic.rnu.tn, Mohamed.jemni@fst.rnu.tn

Abstract. This paper describes the development of a 3D continuous sign language recognition system. Since many systems like WebSign[1], Vsigns[2] and eSign[3] are using Web3D standards to generate 3D signing avatars, 3D signed sentences are becoming common. Hidden Markov Models is the most used method to recognize sign language from video-based scenes, but in our case, since we are dealing with well formatted 3D scenes based on H-anim and X3D standards, Hidden Markov Models (HMM) is a too costly double stochastic process. We present a novel approach for sign language recognition using Longest Common Subsequence method. Our recognition experiments were based on a 500 signs lexicon and reach 99 % of accuracy.

Keywords: Sign Language, X3D/VRML, Gesture recognition, Web3D scenes, H-Anim, Virtual reality.

1 Introduction

Sign language is the natural language of deaf. It is a visual-gestural language which makes it different from spoken language but serves the same function. Sign language is the main way for deaf to communicate, in this context; many projects have been developed to make information accessible to deaf. We can mention WebSign[1] which can generate automatically a 3D signing avatar from a written text. Websign is developed by the research laboratory of technologies of information and communication (UTIC).

Sign language enables deaf people to communicate through a visual receive channel and a gestural transmission channel which makes this language able to transmit simultaneous information. Moreover, Cuxac [4] showed that Sign Language (SL) uses iconicity which enables deaf to describe objects or actions while signing. SL is not a universal language; many different languages have been involved regionally as, e.g., FSL (French Sign Language) in France or ASL (American Sign Language) in the United States. To sign a word, five parameters might be combined. These are hand shape, hand orientation, location, hand motion and facial expression.

Grammar of sign language is different from spoken language. The structure of a spoken sentence is linear, one word followed by another. Whereas in sign language,

simultaneous structures regularly exist. At the syntactic level, sign language is not as strict as spoken language and the OSV¹ order is the most commonly used order.

This paper presents a novel approach to recognize Sign Language from Web3D based scenes. Our method is an adaptation of Longest Common Subsequence (LCS) algorithm which is commonly used for approximate string matching purposes. Our experiments showed that our systems reach up to 99% of accuracy using a dictionary of 500 signed words.

The paper is organized as follows. Section 2 summarizes the aspects of signing avatars systems. In section 3 we review the main difficulties encountered during the recognition process and we illustrate the architecture of our recognition system. In section 4, we outline a summary of what has been done in the field of sign language recognition from video-based scenes. Section 5 presents the main contribution of this paper which consists on the adaptation of the LCS algorithm to recognize sign language from Web3D based scenes. Experimental results are presented in section 6. Finally, the section 7 brings out the conclusion.

2 Signing Avatars Systems

Since the standardization of 3D signing avatars modeling by Web3D consortium[5], many systems have carried out the automatic Sign Language generation from a written text. Web3D Signing avatars are modeled using the H-Anim [6] working group specifications that give a relevant method to describe the skeleton architecture and hierarchy. The animation of these humanoids is rendered according to the X3D specifications [7] that implement a high layer on the VRML (Virtual Reality Markup Language) code.

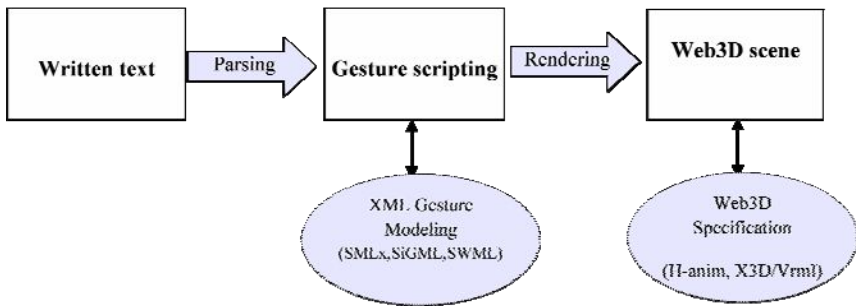


Fig. 1. Architecture of a signing avatar system

3 Related Difficulties and Proposed System Architecture

Although our system is not video-based, we encounter almost the same difficulties. Some of them are related to Sign language itself and others are related to the

¹ Object Subject Verb : common order used in sign language.

technology used to generate signs. Here are the most relevant problems that we encountered during the recognition process:

- 3D Scenes to be parsed may contain formatting mistakes caused by failures to meet H-anim and X3D standards.
- In sign language, for the same signed word, we can find different rotation values for a given join².
- In sign language there are no explicit indications for the beginning and the end of gesture, indeed, the gesture segmentation is the most difficult task that we should consider.

The figure 2 shows the architecture of our system which is based on 3 modules. Every module carries on a part of the recognition process to lead finally to a written text recognized from a 3D scene. The recognition is based on an X3D signed dictionary.

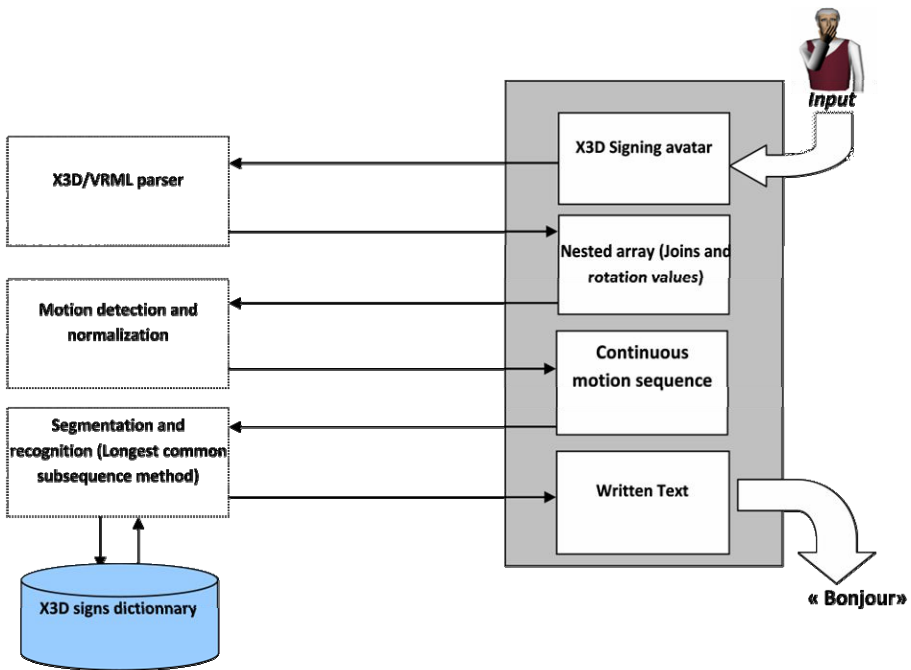


Fig. 2. Proposed system architecture

4 Related Work in the Field of Sign Language Recognition

Sign language recognition has been attempted since several years [15]. In 1997 Vogler and Metaxas [8] described a system for Continuous ASL recognition based on 53-Sign vocabulary. Vogler and Metaxas performed two experiments, both with 97 test sentences: one, without grammar and another with incorporated bi-gram

² Join or articulation of a part of the body ex: r_shoulder.

probabilities. Recognition accuracy ranges from 92.1% up to 95.8% depending in the grammar used. In 1998, Starner [9] proposed a HMM-based recognition system. It recognizes 40 words with 91.3% accuracy.

These video-based systems used signal and image processing techniques to match and tokenize signs like HMM (Hidden Markov Models) which is a double stochastic process based on transition matrix and probabilities. The recognition process can be performed with a Baum-Welch algorithm or a Viterbi algorithm as described by Rabiner [10].

In our case, as WebSign dictionary is composed by more than 1200 signed words, probabilistic methods like HMM become a costly task. Moreover, X3D signs are well formatted and form a rich alphabet. Thus, we are no longer dealing with simple states but with 3D rotation values. In this context, the originality of our contribution consists on using string matching and patterns recognition methods, especially an adapted version of the Longest Common Subsequence algorithm (LCS).

5 Sign Language Recognition Using the LCS Method

5.1 Motion Detection Algorithm

The following algorithm allows us to extract all information about animated joints in a Web3D scene about over time. An example of result is shown in table 1.

```

Program Fetch_Rotations (X3D scene)
  For each (OrientationInterpolator)
    For each (TimeSensor)
      Find Corresponding ROUTE
      RotationsArray[joint]<- ROUTE[toNode]
      For each TimeSlot T
        RotationArray[Time]<- T
        RotationArray[Value]<- KeyValue
      End For
    End For
  End For
Return RotationArray
End.
```

The above algorithm runs after validation of a scene, the table below shows a portion of a returned array by our algorithm. For every time slot, we obtain all rotation values of the avatar's animated joints. To represent 3D rotations, we adopt the representation described in [11] based on quaternion (x,y,z,w).

Table 1. Extracted motion sequence using Fetch_Rotation Algorithm

0.5	JointName	r_elbow			
	RotationValue	X	Y	Z	W
		-0.498537	1.092814	-0.108278	1.955234
	JointName	r_index1			
	RotationValue	X	Y	Z	W
		-12.37785	0.000000	0.000000	0.161756

5.2 Sign Language Segmentation and LCS Problem

The extracted data represents a sequence of 3D keyframes where we have to find out the signed words using a dictionary. Known that scenes to be processed are issued from different systems, the number of keyframes and their rotation values could be different from those recorded in our dictionary. Thus, our problem is similar to the Longest Common Subsequence (LCS) problem [12] where:

- The scene to be processed is the Text: T[KeyFrame₁,...,KeyFrame_n], where keyframes are the returned 3D rotation values by our Fetch_Rotations algorithm.
- The Signed word (recorded in the dictionary) to search in the scene is the pattern : P[KeyFrame₁,...,KeyFrame_m]
- Every KeyFrame belongs to Σ alphabet . $\Sigma = \{r_{elbow}, l_{elbow}, r_{wrist}, l_{wrist}, \dots\}$

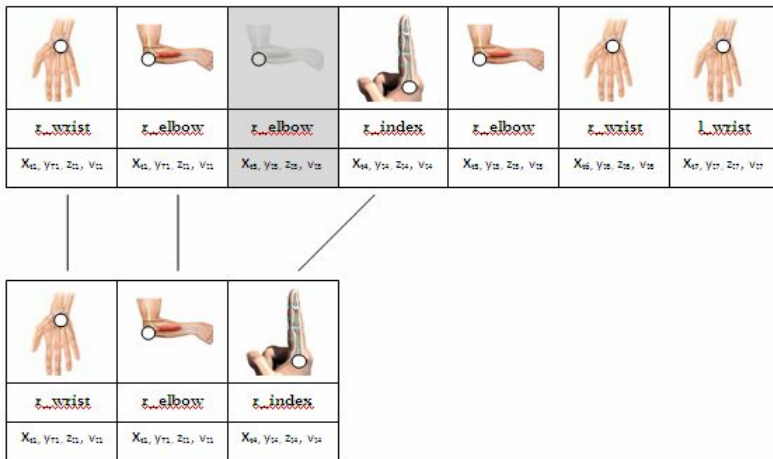


Fig. 3. Adapting the LCS problem to Sign Language

5.3 Multi-dimensional Scaling for 3D Key-Frames Comparison

Our approach is based on dynamic programming. This paradigm needs to find a way how to compare 3D keyframes because we no longer compare simple alphabet letters.

For this purpose, we used Multidimensional Scaling[13] to quantify similarities between 3D keyframes precisely a generalized form of Minkowski distance of order 4 [14] . This allows us to compute the distance between a KeyFrame1 and a KeyFrame2 using the formula below.

$$\text{Minkowski distance of order 4} = \left(\sum_{i=1}^n |p_i - q_i|^4 \right)^{\frac{1}{4}}$$

After calculating the Minkowski distance we have just to convert it into a similarity rate called MinkowskiSimilarity with the formula below :

$$\text{MinkowskiSimilarity} = \text{dmax} - \text{dmax} / \text{dmax} \text{ Where } \text{dmax} = (4 * \pi * 4) / 14 = 8.881$$

For example, if use the formula above to compare the two following keyframes,

- Keyframe1 (from the scene) : l_shoulder (3.00, 2.00, 00.00, 01.00)
- Keyframe2 (from the signed word) : l_shoulder (3.00, 2.50, 99.00, 01.20)

We obtain,

$$\text{MinkowskySimilarity} = (8.881 - 0.7) / 8.881 = 0.92118$$

The rate of similarity between two keyframes, can be used as a threshold beyond which the keyframes are considered as similar. Hence, we use this rate to adapt the LCS algorithm to sign language recognition.

5.4 Adapting the LCS Algorithm

In this section we adapt the Longest Common Subsequence method. X is a signed word recorded in the dictionary and Y is the scene that contains a signed sentence. We apply the dynamic programming to find out the LCS(X, Y). If the length of returned LCS is equal to the length of X then the word exists in the scene.

- X=[r_wrist(0 0 0 1), r_elbow(0.5 0 0 0.5), r_index(1 1 1 1)]
- Y= [r_wrist(0 0 0 1), r_elbow(0.5 0 0 0.5), r_elbow(0 0 1 0), r_index(1 1 1 1), r_elbow(0.1 0 0 -1.3), r_wrist(0 0 0 0) , l_wrist(1.5 0 0 1)].

To solve the LCS problem we proceed in two steps as follows:

1. Distance matrix computing using the following adapted formula :

$$T[i, j] = \begin{cases} 0 & \text{if } i = 0 \text{ or } j = 0 \\ T[i - 1, j - 1] + 1 & \text{if } i > 0 \text{ and } j > 0 \text{ and Similarity}(x_i, y_j) \geq \text{threshold} \\ \max \{T[i - 1, j], c[i, j - 1]\} & \text{if } i > 0 \text{ and } j > 0 \text{ and Similarity}(x_i, y_j) < \text{threshold} \end{cases}$$

2. The second step consists on performing a back-trace to find the LCS(X,Y). We adapted the backtrace algorithm so that it can use similarity as well. Te algorithm takes the scene Y of length n, the pattern X of length n and the distance matrix T computed in the previous step. The table 2 shows the way our algorithm works on the example given previously where the pattern r_wrist, r_elbow, r_index was indeed found.

```

Program TraceBack_Modified (x,m,y,n,T)
    i ← m-1; j ← n-1; k ← T[m-1, n-1] - 1
    While i>0 and j>0 Do
        IF T[i, j] = T[i-1, j-1]+1 && MinkowskiSimilarity(xi, yj) >= threshold
            wk ← xi;
            i ← i-1; j ← j-1 ; k ← k-1
        Else IF T[i-1, j] > T[i, j-1] Then i ← i-1 Else
            j ← j-1
    Return w
    
```

Table 2. Illustration of the Back-Trace algorithm

			-1	0	1	2	3	4	5	6
				ϵ_{wrist}	ϵ_{elbow}	ϵ_{elbow}	ϵ_{index}	ϵ_{elbow}	ϵ_{wrist}	l_{wrist}
				0001	0.5000.5	0010	1111	0.1001.3	0000	1.5001
-1			0	0	0	0	0	0	0	0
0	ϵ_{wrist}	0001	0	1	1	1	1	1	1	1
1	ϵ_{elbow}	0.5000.5	0	1	2	2	2	2	2	2
2	ϵ_{index}	1111	0	1	2	2	3	3	3	3

6 Experimental Results

Our system was tested using a 500 signed words dictionary. The figure 4 shows the recognition performance. The first curve shows that the recognition process takes quite a long time according to the length of the used dictionary and the length of the scene as well. The second curve shows that recognition accuracy varies according to

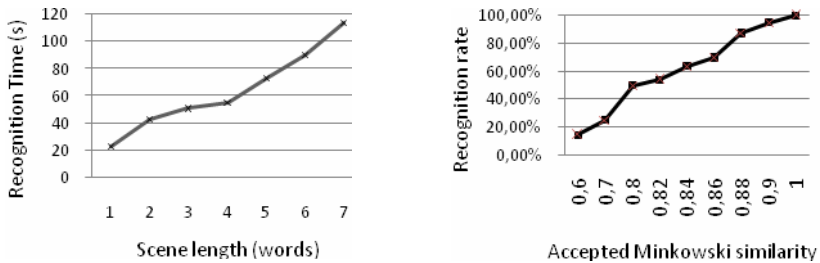


Fig. 4. Performance of the proposed system

the threshold of accepted similarity between keyframes. As shown in the second curve, we can reach 100 % accuracy if the scenes we are processing are generated from our own dictionary. Tests have been performed on a machine with 1Gb of R.A.M and a 1,73 Ghz processor.

7 Conclusion

In this paper, we brought out the fact that signing avatars systems like are becoming increasingly common. We came up with a novel approach to recognize sign language from scenes generated with this kind of systems. We built a system able to detect motion and recognize signed words inside the 3D scenes. Our method is an adaptation of the LCS algorithm commonly used for approximate string matching problems. Experimental results were very encouraging since we reach 99% of accuracy, even if the execution time is an aspect that we should work on in the future.

References

1. Jemni, M., Elghoul, O.: Towards Web-Based automatic interpretation of written text to Sign Language. In: ICTA 2007, Hammamet, Tunisia, April 2007, pp. 12–14 (2007)
2. Papadogiorgaki, M., Grammalidis, N., Makris, L., Sarris, N., Strintzis, M.G.: Vsign Project. Communication (September 20, 2002), <http://vsign.nl/EN/vsignEN.htm>
3. Ehrhardt, U., Davies, B.: A good introduction to the work of the eSIGN project. eSIGN Deliverable D7-2 (August 2004)
4. Cuxac, C.: La LSF, les voies de l'iconicité. Ophrys editions, Paris (2000)
5. Web3D consortium website, <http://www.web3d.org>
6. Humanoid Animation Standard Group, Specification for a Standard Humanoid: H-Anim 1.1, <http://www.h-anim.org/Specifications/H-Anim1.1/>
7. Brutzman, D., Ardly, L.: X3D Extensible 3D Graphics for Web Authors. Elsevier, Amsterdam (2007)
8. Vogler, C., Metaxas, D.: Adapting hidden Markov models for ASL recognition by using three-dimensional computer vision methods. In: IEEE International Conference on Computational Cybernetics and Simulation, Orlando, FL (October 1997)
9. Starner, T., Weaver, J., Pentland, A.: Real-time American Sign Language Recognition using desk and wearable Computer based Video. J. IEEE Transactions on Pattern Analysis and Machine Intelligence (1998)
10. Rabiner, L.: A tutorial on Hidden Markov Models and Selected Applications in Speech Recognition. Proc. Of the IEEE 77, 257–289 (1989)
11. Euclidean Space by Martin John Baker, <http://www.euclideanspace.com/maths/geometry/rotations/conversions/quaternionToEuler/>
12. Bergroth, L., Hakonen, H., Raita, T.: A Survey of Longest Common Subsequence Algorithms. J. SPIRE, 39–48 (2000)
13. Torgeson, W.S.: Multidimensional scaling of similarity. J. Psychometrika 379–393 (2006)
14. Deza, M., Deza, E.: Dictionary of Distances, Elsevier editions, Amsterdam (2006)
15. Jaballah, K., Jemni, M.: Automatic Sign Language Recognition using X3D/VRML Animated Humanoids. In: CVHI 2009, Wroclaw, Poland (April 2009)

Introducing Arabic Sign Language for Mobile Phones

Hend S. Al-Khalifa

Information Technology Department
College of Computer and Information Sciences, King Saud University
Riyadh, Saudi Arabia
hend@ccis.edu.sa

Abstract. With the wide spread of smart mobile phones equipped with advanced processing power among deaf people, the need for utilizing these devices to serve the deaf people requirements is becoming an evolving research trend. In this paper we present the idea of a portable Arabic sign language translator for mobile phones to convert typed Arabic text to sign language animation using 3D signer avatar. This will empower people with hearing disability by using their mobile phones to bridge the divide between them and the hearing people.

Keywords: Sign Language, Arabic Language, Mobile phones, Hearing Impaired, Usability.

1 Introduction

Deaf and hard of hearing people constitute approximately 70 million worldwide and around 17 millions in the Arab world¹. The biggest challenge that faces those deaf people on a daily basis is communication. This makes sign language a very important communication media for them. Sign language is "a language based on movement of hands, arms, body, and facial expressions to create gestures that communicate a person's thoughts" [5].

For Arab people with hearing disabilities, the first, natural, and spontaneity language is the sign language and the written Arabic language is considered a second language for them [1]. This fact led to reading and writing deficiencies among the Arab deaf community. Some deaf people are facing difficulties in understanding the written text and in writing well formed sentences in their own. When a deaf person tries to read Arabic text, (s)he finds difficulties in understanding it [1]. To overcome this problem, technology can be put into use to bridge the communication gap between deaf and non-deaf people.

With the advent of smart mobile phones equipped with powerful processing power similar somehow to personal computers capabilities, many people in the deaf community use these devices for basic communication matters such as sending text messages and e-mails. However, these small and powerful devices can be further utilized to function as a portable translation machines.

In this paper we introduce the idea of a portable Arabic sign language program for mobile phones that converts Arabic text to Arabic sign language. One benefit of using

¹ World Federation of the Deaf, <http://www.wfdeaf.org/>

such a program is to help the Arab deaf community break communication barriers with the outside world and to have a translation program with them anytime and anywhere.

Our translation program is specialized in text to sign language translation only, where text is inputted in the form of formal Arabic language and then translated to unified Arabic sign language that adheres to the Arabic Dictionary of Gestures for the Deaf.

This paper begins by shedding the light on some endeavors for utilizing mobile phones for sign language translation. It will then go on presenting the implementation and evaluation of our translation program. Finally, the paper concludes with highlighting our program limitations and future remarks.

2 Related Work

There are communication barriers between the deaf and the hearing world. From our understanding of such problems, many projects have been implemented to recruit technology to meet the needs of deaf people. These projects have different goals, some were educational to help in learning deaf sign language, and others were mediators for translating text and speech into sign language and vice versa. However, most of these systems were PC-based, little have been developed to leverage the power of the new breed of mobile phones.

For example, a recent research project called "the Sign Language MMS to Make Cell Phones Accessible to the Deaf and Hard-of-hearing Community" has been carried out in Tunisia in 2007 [3]. This project describes a new software module which allows automatic transformation of a textual message to an MMS containing the translation of the message's content to sign language in order to communicate with deaf people via cell phones. This transformation happens in two steps: first, a generation of 3D animation of an avatar, by the use of the WebSign web service. The generated 3D animation is transformed by MMSSign server to an MMS containing a 3GP video sequence which sends it to the server of the telecom operator [3].

The disadvantage of this project is that the translation system resides on a server not on the mobile phone. Another disadvantage is that the receiver does not see the original translated text on his/her mobile phone; this will not help the hearing community who are interested in learning the Sign Language. Finally, the system is costly besides it will not be effective in low bandwidth networks.

Similarly, Halawani [2] proposed the implementation of an Arabic Sign Language Translation System on Mobile Devices. The proposed system consist of four main parts: 1) Mobile device player that allows the user to input the text and to view the animation, 2) the Text translator tool that supports the translation of text in each phase of the communication process, 3) the Matching tool that establish comparison and matching of text and sign, and 4) the sign language that is stored in the Sign Animation Database.

The main difference between Halawani's system and our proposed system is in the connection to a web server. Halawani's system needs a connection to a web server to operate while our system does not need this connection [2].

On the other hand, López et al. [4] described in their paper a special format for defining and storing sign language symbols for a virtual 3D human avatar over mobile

devices. This new format has been designed to be used in mobile devices with low storage capacity and low graphic resources. Yet, no evaluations were undertaken by the research team to prove the effectiveness of their technique.

Finally, Ahmed and Seong [5] have tested the feasibility of writing and reading text messages in sign language as an alternative communication mean besides SMS. The empirical evaluation results affirmed that the test subjects preferred Sign Writing against SMS as the communication mean. In our program we will not use Sign writing for translation; instead we will rely on inputted Arabic text and let our program translate it to sign language.

Given the previously mentioned research, our proposed program will overcome the limitations of [2] and [3] and benefit from the idea presented in [4]. Next we will detail our system architectural design and its implementation.

3 System Overview

Our system was developed using J2ME language, a stripped-down version of Java, suitable for small devices with limited capabilities.

The system consists of two processes, namely: translation and presentation, and four resources files, to act as a dictionary for the translation (Fig. 1).

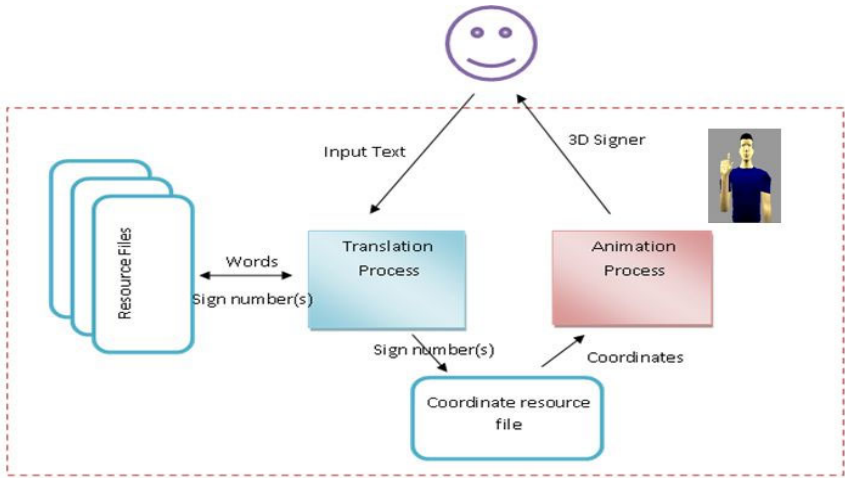


Fig. 1. System Architecture of the Portable Arabic Sign language Program for Mobile Phones

3.1 Translation Process

In the translation process we used the sliding window technique to handle the occurrence of compound words. When translating from Arabic text to Arabic sign language three cases are encountered: compound words which correspond to compound sign representation, single words which correspond to one or more sign language representation and single words that do not have a corresponding sign gesture such as proper names, in this case they can be finger spelled.

A sliding window search is performed by sliding a window of length two words along the text then locating these words in the compound resource file. If the compound word was found, the two words are inserted in a processed text queue as one object and the window is moved to the word after the compound. However if the compound was not found, the first word is classified as a simple word and inserted in the processed text queue, then the window is moved to the second word to be the first word in the sliding window and so on.

In order to deal with the morphological nature of the Arabic language, the translation process uses the general Arabic grammar rules to turn an Arabic noun or verb to its root. The morphology process takes into consideration the mobile phones space and power limitations and at the same time provides a sufficient way to analyze the inputted text by using if-statements with immediate return-statements once the word was found. Therefore, our technique used to handle Arabic grammar will require less space and less processing time than using readymade grammar analysis tools that require very large space for their embedded databases which will not be sufficient to use on mobile phones.

3.2 Presentation and Animation Process

The second process is the presentation and animation process. We used the coordinate technique to animate the 3D signer avatar via the Mobile 3D Graphics API (M3G) to minimize the memory space required for the animation process. This technique is appropriate for mobile memory constraints.

As for storing the words with their corresponding sign movements, we used resource files as a replacement for databases. Resource files are Unicode safe and take small space compared to database systems. Also, reading and dealing with resource files by MIDP (Mobile Information Device Profile) applications is easier and faster than dealing with text files, which helps us to gain more optimization.

In our system we created four resource files: *stem*, *compound*, *character*, and *movement files*. The stem file contains a list of words, the compound file contains a list of compound words, the character file contains a list of letters and numbers, and the movements file contains the movement information associated with each sign number. Each stem, compound, and character has a sign number to relate it with the corresponding movements. Each sign number is equivalent to the word's order in the resource files.

To further optimize the space requirements of our application, the user interface (UI) includes only the basic options a user would need (Fig. 2), that is:

1. **Translation** [ترجمة]: It translates a single word/sentence from Arabic text to Arabic sign language.
2. **Replay** [إعادة]: It replays the sign language translation animation.
3. **Slow Replay** [إعادة بطيئة]: It replays the animation in a slower speed.
4. **Return** [رجوع]: It enables the user to go back to the previous window.
5. **Return to main window** [الرجوع للقائمة الرئيسية]: It enables the user to go back to the main window.
6. **Spelling** [تهجئة]: Finger spell Arabic text to Arabic sign language.
7. **The dictionary** [القاموس]: The user can browse the content of the resource files to lookup the sign gesture of a single word or a compound word or a letter.



Fig. 2. On the left the UI of our program, on the right the 3D signer Avatar representing the first alphabet of the Arabic language

4 Evaluation

To grade the program usability and measure the user satisfaction, we assigned a group of users some tasks to evaluate our program. We then used the System Usability Scale (SUS) which is a Likert scale-based questionnaire made to grade the usability of systems and its results are used to calculate a score between 0 and 100, where 100 indicating “best” usability [6].

Table 1. SUS Survey

Items in the SUS survey
1. I think that I would like to use this system frequently.
2. I found the system unnecessarily complex.
3. I thought the system was easy to use.
4. I think that I would need the support of a technical person to be able to use this system.
5. I found the various functions in this system were well integrated.
6. I thought there was too much inconsistency in this system.
7. I would imagine that most people would learn to use this system very quickly.
8. I found the system very cumbersome to use.
9. I felt very confident using the system.
10. I needed to learn a lot of things before I could get going with this system.

The SUS survey is a ten-item scale which gives a global view of the subjective assessments of the system usability. These items are shown in Table 1.

The SUS survey was distributed among five users consisting of deaf and non-deaf people. They were asked to rate each item from 0 to 4, where 0 indicates strongly disagree and 4 indicates strongly agree.

Table 2 below shows each user total score and the average of their scores.

Table 2. SUS Scores

User ID	User status	User Score
User#1	Deaf	97.5
User#2	Deaf	80
User#3	Deaf	95
User#4	Non-deaf	92.5
User#5	Non-deaf	90
Average = 91		

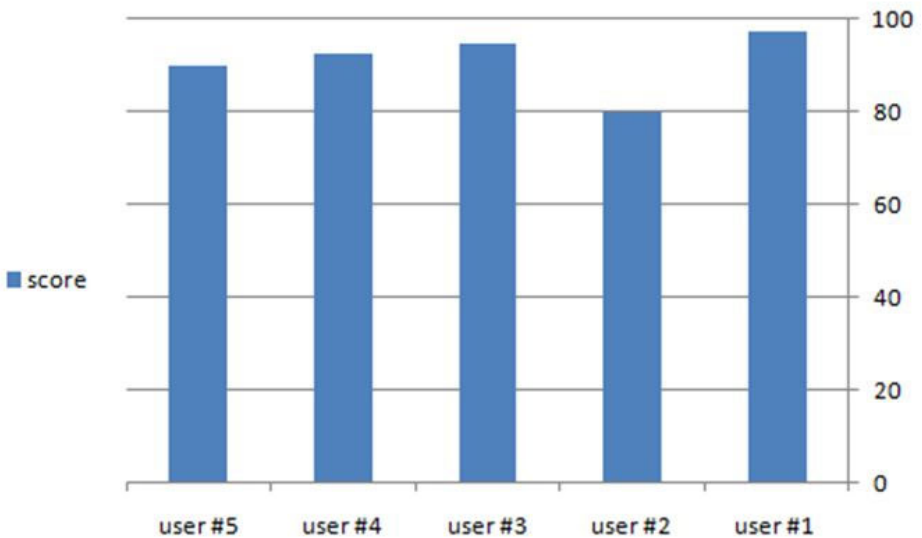


Fig. 3. SUS Scores

From the obtained results we can find that our system has a high SUS score, which indicate a high user satisfaction toward using our program. However the disadvantage of our program that was commented by the test users was that, the program did not show the facial expressions and body motions, which are very important in the representation of some of the words.

5 Conclusion and Future Work

In this paper we presented the implementation and evaluation of a novel mobile phone application that translates Arabic text into unified Arabic Sign Language. The main users of our system are people with hearing disabilities besides non-deaf people who are interested in learning Arabic sign language.

In order to translate Arabic text to Arabic sign language correctly, the entered text should be well formatted and free of syntactical errors.

The avatar that interprets the Arabic Sign Language is created using three dimensional mesh to give the user the impression of a human-like signer.

Even though our system has achieved a high score in System Usability Scale (SUS) evaluation, yet, the limitation of our system can be listed as follows:

- Limited number of characters allowed in a translated sentence (only 50 characters),
- Limited number of words in the dictionaries (resource files), and
- The translation processes do not take sentence semantics into consideration; in fact this is one of the open research topics in Arabic natural language processing.

The future improvements of our system will include: expanding the dictionary to include more Arabic words, and representing face and body expressions.

Finally, it is worth mentioning that our software works only on devices that support Java ME APIs, specifically M3G API which is JSR 184 specification.

An alpha version of the software can be downloaded from the following link <http://jump.fm/DXWFH>.

Acknowledgment

The author would like to thank her undergraduate project students for developing the system.

References

1. Al-Showaier, H.: Standardizing Arabic Sign Language. In: The proceedings of the 3rd International Conference on Disability & Rehabilitation "Scientific Research in the Field of Disability", Riyadh (2009)
2. Halawani, M.: Arabic sign language translation system on mobile devices. IJCSNS International Journal of Computer Science and Network Security, 8 January (1), 251–256

3. Jemni, M., El Ghouli, O., Ben Yahia, N., Boulares, M.: Sign Language MMS to Make Cell Phones Accessible to the Deaf and Hard-of-hearing Community (2007); Electronic version: <http://ftp.informatik.rwth-aachen.de/Publications/CEUR-WS/Vol-415/paper25.pdf>
4. López, F., Tejedor, J., Garrido, J., Colás, J.: Use of a Hierarchical Skeleton for Spanish Sign Language 3D Representation over Mobile Devices. In: VII Congreso Internacional de Interacción Persona-Ordenador, INTERACCION 2006, Puertollano (Spain), November 2006, pp. 565–568 (2006)
5. Ahmed, A.S., Kuen Seong, D.S.: Sign Writing on mobile phones for the deaf. In: Proceedings of the 3rd International Conference on Mobile Technology, Applications and Systems, Mobility. ACM Press, New York (2006)
6. Meiert, J.: SUS: How to Easily Grade Your Site's Usability (December 27, 2009), <http://meiert.com/en/blog/20091127/sus-how-to-grade/> (accessed on January 27, 2010)

Sign Language Interpreter Module: Accessible Video Retrieval with Subtitles

Primož Kosec¹, Matjaž Debevc¹, and Andreas Holzinger²

¹ University of Maribor, Smetanova 17, SI-2000 Maribor, Slovenia
Faculty of Electrical Engineering and Computer Science

² Medical University Graz, A-8036 Graz, Austria
Institute for Medical Informatics, Statistics & Documentation
Research Unit HCI4MED

{pkosec, matjaz.debevc}@uni-mb.si,
andreas.holzinger@meduni-graz.at

Abstract. In this paper, we introduce a new approach to the integration of sign language on the Web. Written information is presented by a Sign Language Interpreter Module (SLI Module). The improvement in comparison to state-of-the-art solutions on the Web is that our sign language video has a transparent background and is shown over the existing web page. The end user can activate the video playback by clicking on an interactive icon. The mechanism also provides a simplified approach to enable accessibility requirements of existing web pages. In addition, the subtitles are stored externally in the Timed Text Authoring Format (TTAF), which is a candidate for recommendation by the W3C community. Empirical results from our evaluation study showed that the prototype was well accepted and was pleasant to use.

Keywords: Web, Multimedia, Video, Sign Language, Subtitles, Deaf, Hard of hearing, Human-Computer Interaction (HCI), Usability Engineering.

1 Introduction

The World Wide Web (the Web) has evolved to become an interactive multimedia environment. However, information for the broader population is mainly provided as written text, static images or static videos. The existing accessibility guidelines for rich material on the Web (Web Content Accessibility Guidelines, International Telecommunication Union and International Organization for Standardization) are unfortunately too general and inappropriate for the specific needs of users with disabilities, such as deaf people who use sign language as their first language and consider the national language as their second language. The areas of difficulty for deaf people include a possible lack of reading and writing abilities. Based on research studies, many have a low level of literacy and a poor understanding of wording. For this reason, the written information on the Web should alternately be presented with videos of sign language interpreters.

In this paper, based on our extensive research on the needs and requirements of the deaf and hard of hearing community [1], we introduce the next prototype of Sign Language Interpreter Module (SLI Module), which considers the following four aspects:

- (a) Providing accessibility feature for deaf and hard of hearing to existing web pages;
- (b) Inclusion of various form of information (video, audio and subtitles);
- (c) Applying subtitling Timed-Text Authoring Format [22]; and
- (d) Cross-platform browser Flash Player plug-in.

2 Background and Related Work

The existing solutions on the Web, for example SMILE [2], ShowSounds [3], Signing Web [4], ATBG [5], SignOn [6], History of the Deaf [7] and Signing Savvy for American Sign Language [8], use various Flash player plug-ins, which are, in most cases, embedded at a predefined web page location. In this way, web designers have to acknowledge the sign language video location while designing web pages. Existing informative websites, which do not consider accessibility recommendations, frequently contain text, images and photos which expand throughout the whole page, leaving no space for the additional sign language video. The previous work, within other research and projects, suggests that for deaf people who use sign language as their first language, a similar idea has not yet been implemented. Because of its simplicity and appropriateness of usage the solution could be integrated into "WCAG Guidelines" for the needs of deaf and hard of hearing people. Currently, the Web Content Accessibility Guidelines definition states that *"using clear and simple language also benefits people whose first language differs from your own, particularly those people who communicate primarily with sign language"* [9], and does not yet determine that the transparent video should be used, which would solve the problem of rapid and easy integration of sign language videos. The main question arising here is how to provide accessible web pages to deaf and hard of hearing web users with sign language videos and subtitles in an effective way, without redesigning entire website from scratch or drastically changing the layout?

3 The Sign Language Interpreter Module

The initial idea was to provide the deaf and hard of hearing with a new approach for easy and rapid access to information, tailored to their needs, while not discriminating against other users (non-signers). Past research has shown that deaf signer users, who use sign language as a first and desired language, are often helpless and become confused when searching for information on web pages [7][10]. The deaf and hard-of-hearing population is already familiar with a powerful media channel, the television. In comparison to the Web conceptuality, it is more dynamically driven and visual. For deaf and hard of hearing television viewers, the original video and audio is streamed in the background while the sign language interpreter is layered over with additional subtitling. Our proposal for providing accessibility to web pages is the Sign Language Interpreter Module (SLI Module) [11], which presents an addition in WCAG 2.0

(Guideline 1.4 Distinguishable: "Make it easier for users to see and hear content including separating foreground from background" [9]). In this segment, one can find instructions and guidelines for colour, audio, contrast, resize and images of text. However, it does not provide instructions and guidelines for the sign language video, which occurs mainly in Guideline 1.2 Time-based Media: "Provide alternatives for time based media, which is designed to access to time-based and synchronized media on the Web". The SLI Module uses a custom Flash player (SLI Player), which acts as a wrapper of various modalities: a video of a sign language interpreter, audio translations and written translations as subtitles (Fig.1). The video is shown over the web page with a transparent background and video playback controls (a button for hiding/terminating the video is also provided).

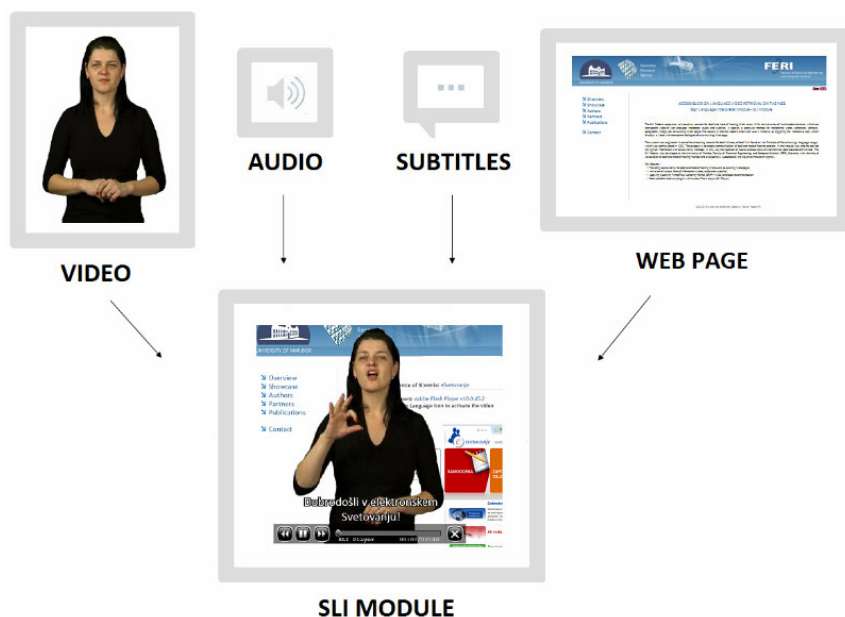


Fig. 1. Inclusions of SLI Module: video, audio, subtitles integrated into a web page

3.1 Video Preparation Phase

The starting phase is the video and audio preparation. The deaf and hard of hearing who use sign language require the best possible video quality in order to focus on the details, such as finger movements for the sign language and lip movements for lip-reading. Therefore, the quality of the video can only be assured by using a high quality digital video camera. The video of the sign language interpreter should be recorded in front of a green background. There are some regulations for interpreter's clothing (for example, black clothes) which provide the contrast between the person standing and the background. The lighting of the recorded subject is also of great importance. The video recorded on the computer must be uncompressed and in original size. For multimedia container we have used the uncompressed Audio Video

Interleave (AVI), with a resolution of 750 x 567 and 25 frames per second (fps), with 48 kHz audio and 32-bit sampling. The video was then imported into the modelling software, where the Color Key effect was used to remove the green background and soften the edges between the object and the background. This procedure and the use of the Shockwave Flash (SWF) format has resulted in the high quality of the video, presenting the moves of the interpreter clearly enough for the users to see the facial mimicry and to focus on the fingers without a blurry image.

In post-production of the video preparation phase, the result appears as video with a transparent background (non-opaque). The video is then converted to a Flash .FLV format. The audio of a qualified person reading the transcript was recorded additionally.

3.2 Custom Flash Player as a Wrapper

Adobe reported that the Flash Player plug-in was installed on 99% of Internet-enabled desktop computers worldwide [12]. There are not many Flash players on the market that support TTAF. JW FLV Player [13], which is a free solution for non-commercial purposes and the ccPlayer [14] were initially investigated. Some of the main reasons why we developed a custom Flash player included implementation-flexibility (layering approach and customization of user interface) and proprietary-independence, which were crucial points for further improvement of the graphical user interface based on evaluation studies with deaf and hard of hearing participants. ccPlayer uses a component called CCforFlash for displaying captions with Flash video and audio content. These captions are stored in external files formatted in the TTAF. The SLI Player uses the Adobe Flash FLVPlaybackCaptioning component (version 2.5) for handling subtitles. The component supports Timed-Text (TT) tags for captioning XML files. The style of font and timing are defined by predefined tags. The limitation is the text outline attribute (*tt:textOutline*), which is used to “*specify a style property that defines a text outline effect to apply to glyphs that are selected for glyph areas generated by content flowed into a region*”. To date, this functionality has been enlisted as a non-supported attribute by Adobe Flash Live Docs [15]. The same effect of text outline can be accomplished through the programming language ActionScript 3.0, which is based on ECMAScript, the international standardized programming language for scripting. SLI Player uses this approach and correctly renders the subtitles' appearance; fonts are displayed as white, surrounded by black outline (like on television).

3.2 Authoring Timed-Text Subtitles

There are not yet many free tools for authoring subtitles in the TTAF. The present online solutions did not meet our requirements for authoring subtitles. For example, YouTube Captions [16], Overstream [17] and DotSub [18] provide subtitling solutions but with the restriction that the videos are hosted on their web servers. One of the online authoring tools that does not need hosted videos is the SubtitleHorse [19], which is a Flash-based subtitle editor, generating subtitles in various formats, including TT. There are also a few offline applications. Media Access Generator (MAGpie) [20] is a free closed-captioning tool, which is designed to produce and synchronize information for QuickTime™ player. Since the latest version 2.5 of MAGpie does not

support FLV video types, and we found the usage was unfriendly and unnecessarily complicated (no automated process), we investigated another free authoring tool. The Subtitle Workshop [21] from UruWorks is an application for synchronizing subtitles with the video and generating subtitles in all kind of formats. It provides playback FLV video types. The authoring process is quite simple. Although the TTAF export feature is enabled, it does not work in the present version (4.0 beta). For this reason, we have implemented the SLI Parser, which is an automated PHP script for converting .SRT subtitling formats to .XML TTAF. The SLI Parser adds also styling-tags for the outline effect, described in Section 3.2. The workflow of producing the TTAF subtitles is presented in Fig. 2.

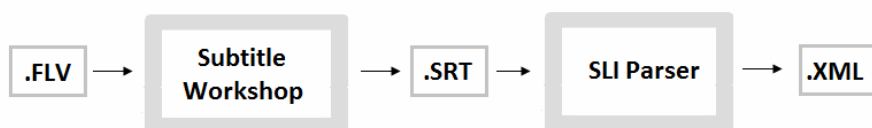


Fig. 2. Workflow of TTAF. The .FLV video is imported into Subtitle Workshop where the synchronization process with the video playback is done. The output file is the .SRT format and contains time-stamps with text. Next, the SLI Parser converts this file to compliant TTAF.

4 The Study

4.1 Participants

Thirty-one deaf and hard of hearing students (age range 15-21 with an average age of 17.8 years) from the Ljubljana School for the Deaf and hard of hearing participated in our study to evaluate the prototype. Seven of the thirty-one pupils were in elementary school and twenty of them were in secondary school. There were 23 males and 8 females. 11 were deaf, 14 were hard of hearing, 5 had a speech disorder and one with autism. 14 of them rated their sign language proficiency as excellent, 12 as good and 5 as poor. 18 participants used the Internet daily, 12 occasionally and one never. A user and task analysis was conducted, where data were collected using simultaneous verbal translations with sign language interpreters, note taking and pre/post study questionnaires.

4.2 Experimental Design

The aims of the experiment were to examine the end-users feedback of the prototype. It was also used to evaluate satisfaction, acceptance and ease of use among deaf and hard of hearing users. The test subjects were gathered in a controlled environment where the prototype was applied - a mock-up of a real website scenario was designed. Participants were shown a demonstration on the purposes and the various usages and functionalities before the experiment. The test subjects were requested to examine a video (5 minutes long), and as reinforcement training, they were allowed to explore the system alone for an additional 10 minutes. At the end of the session, there was a questionnaire about video preferences. Participants were asked to rank the quality of

their experiences using a 5-point Likert scale ranging from »strongly disagree« to »strongly agree«. Areas of evaluation were quality of video and subtitles, media navigation features and parameters for general impressions, satisfaction, ease of use and comprehension. Finally, there was an open discussion with the help of a sign language interpreter. The purpose of an open debate was to encourage participants to reveal their positive and negative effects or missing features of the prototype.

4.3 Results and Discussion

The results reported here relate to participant responses on various aspects of the SLI Module, and are summaries of the notes taken during the study, post-study questionnaires and a group debate. Percentage results of post-test questionnaires are shown in Table 1. 77.3% of the participants only watched the hand movement, however; this was influenced by the high percentage of ‘undecided’ lip-readers (32.2%), who do not generally use lip-reading in everyday life. Subtitles were big enough 77.8% (22 pixels were used) and readable for 77% participants. The video playback navigation controls, were useful for 77.4% of the subjects, while the size was a bit questionable. Those who disagreed (12.9%) suggested smaller buttons. Participants were highly satisfied (80.6 %) with the prototype, 77.4% think the usage is simple. One of the most significant indicators of the evaluation results was the comprehension, which was ranked highly. Due to different modalities (visual, audible or textual) users can choose their modality according to their needs (in some cases all three but in most sign language with subtitles).

Table 1. Post-questionnaire results (%)

		Strongly Disagree or Disagree	Undecided	Agree or Strongly Agree
Sign Language Interpreter	Lip-reading	19.4	32.2	48.3
	Hands-movement	3.2	19.3	77.3
Subtitles	Size	9.7	12.9	77.8
	Readability	6.5	16.1	77.0
Video playback controls	Size of buttons	12.9	19.3	67.7
	Usefulness	12.9	9.7	77.4
Usability	Satisfaction	6.4	12.9	80.6
	Ease of use	6.5	16.1	77.4
	Comprehension	3.2	12.9	83.8

The group debate confirmed some important implications for the further development of the prototype. Participants claimed that the video was too fast, and that they would need to slowdown the video, hence it is appropriate to implement a feature for slowing the video playback of the sign language interpreter. Few participants wanted to move the video around the screen; one suggested that the video could be shown in the right-bottom corner similar to on television. Almost all participants were the same

mind: video subtitles and speech must be synchronized without any latency. Therefore, authoring subtitles by a sign language interpreter was suggested by interpreters who were also teachers for deaf and hard of hearing students.

5 Conclusion and Future Outlook

In this paper, we have provided a technical overview of the Sign Language Interpreter Module. The system presents a combination of multimodal information such as video, audio and subtitles, and offers the option of prioritizing the sign language on the Web for deaf and hard of hearing users. The evaluation findings of the second prototype again showed highly motivated test subjects and their great desire to use this form of sign language presentation. Some of the prototype's features, for example slowing the video playback, moving the video around the screen and the redesign of the video playback controls, have to be taken into account for further development. We also found that, although the synchronization latency of video and subtitles was minimal, it was perceived as a distractive factor. Therefore, authoring subtitles has to be done by a qualified person (such as a sign language interpreter). We think that the system is feasible as it essentially changes the way of retrieving information on the Web. The SLI Module is gaining wide acceptance within the deaf and hard of hearing community as it assists them to obtain written information on the Web in various modalities. This would increase their literacy in their second language and enable them to integrate more easily into the social majority, while at the same time preserving their identity.

Acknowledgments. The project was partially supported by the European commission within the framework of the Leonardo da Vinci program, project DEAFVOC 2. It is also partially supported by the Slovenian Research Agency in the framework of the Science to Youth program, financial support for young researchers. Special thanks go to Živa Peljhan from the Ljubljana School for the Deaf and all participants involved in the study, as well as to Milan Rotovnik from University of Maribor.

References

1. Debevc, M., Kosec, P., Holzinger, A.: Improving Multimodal Web Accessibility for Deaf People: Sign Language Interpreter Module (SLIM). *Multimedia Tools and Applications* (2010) (in print)
2. Kronreif, G., Dotter, F., Bergmeister, E., Krammer, K., Hilzensauer, M., Okorn, I., Skant, A., Orter, R., Rezzonico, S., Barreto, B.: SMILE: demonstration of a cognitively oriented solution to the improvement of written language competence of deaf people. In: 7th International Conference on Computers Helping People with Special Needs (ICHP), Vienna, pp. 17–21 (2000)
3. Vanderheiden, G.C.: Full Visual Annotation of Auditorially Presented Information for Users Who Are Deaf: ShowSounds. In: *Proce. of the RESNA International Conference*, Toronto (1992)
4. Fels, D.I., Richards, J., Hardman, J.L., Daniel, G.: Sign Language Web Pages. *American Annals of the Deaf* 151, 423–433 (2006)

5. Straetz, K., Kaibel, A., Raithel, V., Specht, M., Grote, K., Kramer, F.: An e-Learning Environment for Deaf Adults. In: Proc. of the 8th ERCIM Workshop: User Interfaces for All, Vienna (2004)
6. Hilzensauer, M.: Information Technology for Deaf People. LNCS, vol. 19, pp. 183–206. Springer, Heidelberg (2006)
7. Debevc, M., Peljhan, Z.: The role of video technology in on-line lectures for the deaf. *Disability and Rehabilitation* 26, 10548–10598 (2004)
8. Signing Savvy for American Sign Language, <http://www.signingsavvy.com/index.php>
9. Web Content Accessibility Guidelines 2.0, <http://www.w3.org/TR/WCAG20/>
10. Fajardo, I., Cañas, J.J., Salmerón, L., Abascal, J.: Improving deaf users' accessibility in hypertext information retrieval: are graphical interfaces useful for them? *Behaviour & Information Technology* 25, 455–467 (2006)
11. Sign Language Interpreter Module, <http://slimodule.com/>
12. Adobe Flash Player Statistics, http://www.adobe.com/products/player_census/flashplayer/
13. JW FLV Player, <http://www.longtailvideo.com/>
14. ccPlayerhttp, http://ncam.wgbh.org/invent_build/web_multimedia/tools-guidelines/ccplayer
15. Adobe Live Docs, Flash CS3 Documentation, Timed Text Tags, http://livedocs.adobe.com/flash/9.0/main/wwhelp/wwhimpl/common/html/wwhelp.htm?context=LiveDocs_Parts&file=00000611.html
16. YouTube Captions, <http://captiontube.appspot.com/>
17. OverStream, <http://www.overstream.net/>
18. DotSub, <http://dotsub.com/>
19. SubtitleHorse, <http://subtitle-horse.org/>
20. MAGpie, <http://ncam.wgbh.org/webaccess/magpie/index.html>
21. SubtitleWorkshop, <http://www.urusoft.net/products.php?cat=sw>
22. Timed-Text Authoring Format (TTAF), <http://www.w3.org/TR/ttaf1-dfxp/>

Real-Time Walk Light Detection with a Mobile Phone

Volodymyr Ivanchenko, James Coughlan, and Huiying Shen

The Smith-Kettlewell Eye Research Institute, San Francisco CA
{vivanchenko, coughlan, hshen}@ski.org

Abstract. Crossing an urban traffic intersection is one of the most dangerous activities of a blind or visually impaired person's travel. Building on past work by the authors on the issue of proper alignment with the crosswalk, this paper addresses the complementary issue of knowing when it is time to cross. We describe a prototype portable system that alerts the user in real time once the Walk light is illuminated. The system runs as a software application on an off-the-shelf Nokia N95 mobile phone, using computer vision algorithms to analyze video acquired by the built-in camera to determine in real time if a Walk light is currently visible. Once a Walk light is detected, an audio tone is sounded to alert the user. Experiments with a blind volunteer subject at urban traffic intersections demonstrate proof of concept of the system, which successfully alerted the subject when the Walk light appeared.

Keywords: blindness, visual impairment, traffic intersection, pedestrian signals.

1 Introduction

Crossing an urban traffic intersection is one of the most dangerous activities of a blind or visually impaired person's travel. For some pedestrians with poor or non-existent vision, the rhythm of traffic sounds may indicate when it is time to cross; at some intersections, Audible Pedestrian Signals (APS) provide a clearly audible signal at that time. Unfortunately, such signals are either unreliable or absent at many traffic intersections, and may be inaccessible to pedestrians with combined hearing and vision impairments.

Several types of technologies have been developed to assist blind and visually impaired individuals in crossing traffic intersections. Most prevalent among them are Audible Pedestrian Signals, which not only generate sounds that tell blind and visually impaired individuals when to cross intersections, but also furnish acoustic cues that some pedestrians can localize to help them properly orient themselves to the crosswalk [2]. Talking Signs® [5] allow blind travelers to locate and identify landmarks, signs, and facilities of interest, at intersections and other locations, using signals from installed infrared transmitters that are converted to speech by a receiver carried by the traveler.

However, the adoption of both Audible Pedestrian Signals and Talking Signs® is very sparse, and they are completely absent in most US cities. More recently, Bluetooth beacons have been proposed [3] to provide real-time information at intersections

that is accessible to any user with a standard mobile phone, but like Talking Signs® this solution requires special infrastructure to be installed at each intersection.

The alternative approach that we propose in this paper is to use a portable computer vision-based system, named “Crosswatch” [6,7], to identify important features in an intersection, including the Walk light. With this system (Fig. 1), the user takes an image of the intersection with a mobile phone camera, which is analyzed by software run on the phone, and the output of the software is communicated to the user with audio tones. The great strength of the computer vision approach is that it offers virtually unlimited potential for acquiring information about one’s surroundings, but does not require any infrastructure beyond what is already provided for normally sighted people – namely, pedestrian signal lights and painted street markings.



Fig. 1. Overview of proposed Crosswatch system. Left to right: (a) Blind pedestrian holding mobile phone. (b) Image of crosswalk taken from across the street. (c) Zoomed-up area of (b) shows Walk light visible at the end of the crosswalk. Proposed system is programmed to generate audio alert when Walk light is detected.

Experiments with a blind volunteer subject at urban traffic intersections demonstrate proof of concept of the system, which successfully alerted the subject when the Walk light appeared. This is the first example of a truly portable system we are aware of that provides timing information at traffic intersections based solely on existing traffic signals already used by sighted pedestrians, without the need for any added infrastructure (such as APS or other guidance systems requiring on-site installation).

2 Finding Walk Lights in Images

The Walk light is a pedestrian signal that is lit when it is time for pedestrians to cross. Walk lights have different shapes and colors in different places; we focus on a variety of Walk light that is common in the US, a white icon in the shape of a person walking (Fig. 1c). (It should be straightforward to extend our approach to encompass other Walk light shapes and colors.) After the Walk period ends, the Walk light is replaced by a Don’t Walk symbol, indicating when pedestrians should not cross. (At some traffic intersections there is also an intermediate “countdown” timer signal indicating the amount of time remaining before the Don’t Walk symbol appears, but we defer analysis of this type of signal for future work.) Our goal is to design a computer vision algorithm to rapidly detect the Walk light whenever it is present, so that the user is alerted for as long as it is lit.

There is a variety of past work on traffic signal light detection in computer vision, including some research aimed at vehicle traffic light detection [4] and other work [1] specifically addressing Walk light detection for visually impaired pedestrians. All the work that we are familiar with in this vein exploits the fact that the lights being detected are colored (i.e. not white) in order to narrow the search for targets of the appropriate shape; for our application, however, the Walk light is white, so color is not as powerful a cue. Moreover, the algorithms described in this work were all implemented on portable PCs, whereas we seek a solution that runs on the much more convenient (but less computationally powerful) mobile phone platform.

To add to the challenge of a reliable, real-time solution, we note that the Walk light appears very small at typical viewing distances (across the street – see Fig. 1b for a representative example). Moreover, in order to process images at least one or two frames per second, we are forced to acquire the images in video mode, which has much lower resolution (VGA, i.e. 640 x 480 pixels) than the highest-resolution still images available on the Nokia N95 (5 Megapixels, i.e. 2592 x 1944 pixels). As a result, the Walk light appears to be small and blurry (and is sometimes subject to color distortions) in typical images (Fig. 2), which makes recognizing it on the basis of its shape more difficult.



Fig. 2. Zoomed-up images of Walk lights, photographed from typical viewing distances, rendered at the lower resolution available to the camera running in video mode. Note that the Walk icon is blurry, hard to resolve clearly, and may be subject to color distortions.

Our solution to these challenges is to narrow our search using a powerful additional cue: at typical viewing distances, the Walk light is far enough away that it appears close to the horizon. If we can estimate the location of the horizon in the image, then we need only search a narrow strip of the image near the horizon rather than the entire image (Fig. 3a). Fortunately, most modern mobile phones (including the N95) have built-in accelerometers (i.e. tilt sensors) that continuously estimate the direction of gravity [7], which in turn determines the horizon line with only a few simple calculations.

Our search for the Walk icon in the image strip near the horizon proceeds in several stages. First, the image strip is analyzed to determine which pixels are bright and close to white in color, and which pixels are dark (brightness and darkness are assessed relative to other pixels in the strip). Second, candidate Walk icons are

identified by finding small regions of bright whitish pixels surrounded by dark pixels (Fig. 3a). Third, the shape of these candidates is compared with the correct Walk icon shape using template convolution to score the candidates and rule out the ones that have an obviously incorrect shape. Finally, for the top-scoring 15 candidates, the shape is explicitly analyzed for the presence of a two legs and a head, which must be visible (even if the resolution is poor).

The above detection procedure is fast and effective in many intersection scenes, but to further reduce the likelihood of false positive Walk light detections we conduct an additional test: the candidate Walk light must appear near to and below a green traffic light (Fig. 3b), which is detected as a bright green light that is surrounded by dark pixels. While there are some rare traffic intersections where a Walk light is *not* accompanied by a nearby green light, we accepted this as a temporary restriction for our prototype system.

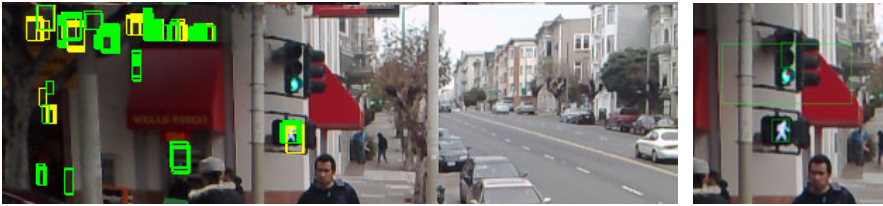


Fig. 3. Left to right: (a) In image strip near horizon, candidate regions are selected for processing by detection algorithm. (b) Final Walk light detection, with confirmation by presence of green traffic light above.

3 User Interface

The preceding Walk light detection algorithm was developed and tested on a desktop computer and then ported to the Nokia N95 camera phone running Symbian C++ in video capture mode. The algorithm was implemented to run as quickly as possible on the cell phone, so that at least one image could be processed per second.

In order for the detection algorithm to work reliably, we decided in our prototype Walk light detection system to require the user to hold the mobile phone camera close to the horizontal. Since many blind and visually impaired users have little or no visual feedback about the angle they are holding the camera at, we added a simple user interface to help them correct for unintentional deviations from the horizontal: any time the camera tilt or slant exceeds a small amount (as determined by the direction of gravity estimated by the accelerometer), the system alerts the user with beeping sounds. With a small amount of practice, users learn how to quickly align the camera to the horizontal in order to stop this beeping and allow the system to work properly. (In the future, we will investigate the possibility of using the tilt/slant estimates to automatically unrotate the images captured by the camera so that they appear horizontal to the detection algorithm.)

4 Experimental Results and User Testing

We explained the purpose and operation of the system to a completely blind volunteer subject, and advised him on how best to hold the camera phone. Particular emphasis was placed on the importance of moving the camera slowly to avoid motion blur, ensuring the camera lens was not covered (e.g. by the user's fingers), and holding the camera close to horizontal. After a brief indoor training session with an image of a Walk light scene displayed on a computer monitor, we took the subject to a nearby traffic intersection to test the system. The subject stood across from three separate Walk lights in the intersection, and during two Walk cycles in front of each Walk light (a total of six trials), the system issued approximately ten "Walk" tones per Walk cycle (each Walk signal lasted either 14 or 23 sec.). No "Walk" tones were issued when the Walk light was not lit, although the experimenters encountered occasional, isolated false positives when testing the system themselves.

The main challenge in using the system was to point the camera approximately in the direction of the Walk light, the risk being that the camera may be pointed too far to the left or right to view the Walk light (even if it is held perfectly horizontal). We experimented with using other Crosswatch functionality previously developed [7] to detect two-stripe crosswalk stripes (the two stripes are on either side of the Walk light, as in Fig. 1b) as a way of pointing the user in the correct direction before searching for Walk light. For our test subject this functionality did not seem to be necessary, but it seems natural to integrate crosswalk and Walk light detection in future versions of the system to make it easier to use.

5 Conclusion

We have demonstrated a novel mobile phone-based computer vision system for helping blind and visually impaired pedestrians detect Walk lights to help them cross traffic intersections. A prototype system has been implemented on the Nokia N95 camera phone, which searches for the Walk light at least once per second and provides audio feedback whenever a crosswalk is detected. We have conducted preliminary experiments with blind volunteers to test the system, demonstrating its feasibility.

Future improvements and extensions to the system will emphasize improving its accuracy (especially eliminating false positives), removing the requirement that a green traffic light be present to confirm the Walk light, and adding the ability to detect Don't Walk lights and countdown timers. We will experiment with using the phone's built-in vibrator to provide tactile feedback for users with hearing impairments (or for users with normal hearing in noisy environments). Porting the software to newer and more powerful mobile phones such as the iPhone and Android will allow us to acquire higher-quality video images and improve the speed of the computer vision algorithms. Extensive user testing will be conducted to guide the development of the system and to optimize its usefulness and usability.

Acknowledgments. The authors were supported by The National Institutes of Health grant 1 R01 EY018345-01.

References

1. Aranda, J., Mares, P.: Visual System to Help Blind People to Cross the Street. In: Miesenberger, K., Klaus, J., Zagler, W.L., Burger, D. (eds.) ICCHP 2004. LNCS, vol. 3118, pp. 454–461. Springer, Heidelberg (2004)
2. Barlow, J.M., Bentzen, B.L., Tabor, L.: Accessible pedestrian signals: Synthesis and guide to best practice. In: National Cooperative Highway Research Program (2003)
3. Bohonos, S., Lee, A., Malik, A., Thai, C., Manduchi, R.: Cellphone Accessible Information via Bluetooth Beaconing for the Visually Impaired. In: Miesenberger, K., Klaus, J., Zagler, W.L., Karshmer, A.I. (eds.) ICCHP 2008. LNCS, vol. 5105, pp. 1117–1121. Springer, Heidelberg (2008)
4. Chung, Y.-C., Wang, J.-M., Chen, S.-W.: A Vision-Based Traffic Light Detection System at Intersections. *Journal of Taiwan Normal University: Mathematics, Science and Technology* 47(1), 67–86 (2002)
5. Crandall, W., Bentzen, B., Myers, L., Brabyn, J.: New orientation and accessibility option for persons with visual impairment: transportation applications for remote infrared audible signage. *Clinical and Experimental Optometry* 84(3), 120–131 (2001)
6. Ivanchenko, V., Coughlan, J., Shen, H.: Crosswatch: a Camera Phone System for Orienting Visually Impaired Pedestrians at Traffic Intersections. In: Miesenberger, K., Klaus, J., Zagler, W.L., Karshmer, A.I. (eds.) ICCHP 2008. LNCS, vol. 5105, pp. 1122–1128. Springer, Heidelberg (2008)
7. Ivanchenko, V., Coughlan, J., Shen, H.: Staying in the Crosswalk: A System for Guiding Visually Impaired Pedestrians at Traffic Intersections. In: Association for the Advancement of Assistive Technology in Europe (AAATE 2009), Florence, Italy (2009)

An Ultrasonic Blind Guidance System for Street Crossings

Satoshi Hashino and Sho Yamada

Tokyo University of Technology, School of Computer Science,
1404-1, Katakura, Hachioji, Tokyo-Metropolis, 192-0982, Japan
hashino@cs.teu.ac.jp

Abstract. This paper addresses the technical feasibility of a guidance system based on ultrasonic sensors to aid visually impaired people to cross a road easily and safely. A computer processes ultrasonic signals emitted by a transmitter, which is carried by the impaired user, and provides real-time information on the direction and distance to keep user on the correct track. Instead of time of flight, the system estimates user position by the order of received ultrasonic signals at multiple receivers. Experimental results are presented to discuss feasibility of this method.

Keywords: guidance system, ultrasonic sensor, cross-correlation.

1 Introduction

There are many researches worldwide looking into many forms of travel aid for visually impaired people. These involve technologies like GPS, sonic or ultrasonic signals and or even advance image processing systems [1][2][3]. In comparison to other technologies many blind guidance systems use ultrasound because of its immunity to the environmental noise. Another reason why ultrasonic is popular is that the technology is relatively inexpensive, and also ultrasound emitters and detectors are small enough to be carried without the need for complex circuitry. Salah et al. use two ultrasonic sensors mounted on the user's shoulders and another one integrated into the white cane to perform obstacle detection. Also there have been many devices developed including the Russell Path sounder, AFBs computerized travel aid, or the MOWAT sensor, some of the first ultrasonic devices to be developed.

In Japan and particularly in Tokyo, bumpy yellow color tiles are installed along sidewalks. It functions as a Braille to guide blind when walking along the street. Typically tiles with round bumps indicate end of the pad or intersection or step, and tiles with longer bumps indicate direction. One major problem however, is that these tiles are not laid on street crossings, which makes it difficult to a blind to cross the street without any help. There exists some audible sounds that inform status of traffic light, but there are no devices to inform correct direction to the user. Current paper looks for the technical feasibility of an ultrasonic system to cover this lack for a safe and easy navigation across the street. In the next section we will discuss about the theory followed by experimental results and conclusions.

There are different combinations in which ultrasound sensors are used. Some use portable transmitters, other use portable receivers, and some other use a combination of both. Transmitter site needs simple system as fix circuit, otherwise receiver site need comparatively complex system as computer. A system in which user carries transmitter and complex system is installed on fix site is adopted by virtue of user system compactness. User also might to carry an FM receiver system for getting position error information transmitted from fix site.

2 The Theory

The basic concept of the system is shown in Fig. 1. The straight line on the center of zebra pattern, starting from point *S* and ending at point *E* is the guidance route. On the opposite side of the street, three ultrasonic receivers *R*₁, *R*₂ and *R*₃ are arranged with equal distance *d* on the axis perpendicular to *SE* line. The transmitter circuit carried by the user at point *T*(*x*,*y*) emits ultrasonic waves, which travel through the air and reach to *R*₁, *R*₂ and *R*₃ at three different times *t*₁, *t*₂ and *t*₃ respectively.

Assuming that the flight speed of ultrasonic waves is equal in all directions, the time of flight *t*₁, *t*₂ and *t*₃ will represent the distance of user at point *T* to three receivers *R*₁, *R*₂ and *R*₃. According to Fig. 1, the distance from user at point *T*(*x*,*y*) to these three receivers, labeled as *d*₁, *d*₂ and *d*₃ respectively, can be represented by:

$$d_1 = r_1 \tag{1}$$

$$d_2 = r_1 + r_2 \tag{2}$$

$$d_3 = r_1 + r_3 \tag{3}$$

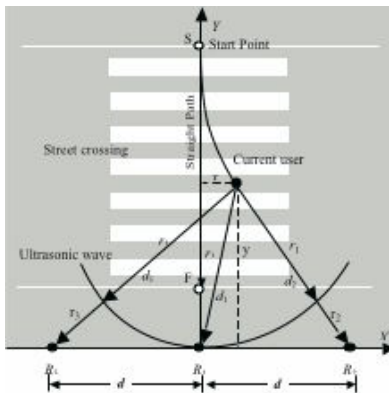


Fig. 1. Concept of the introduced system

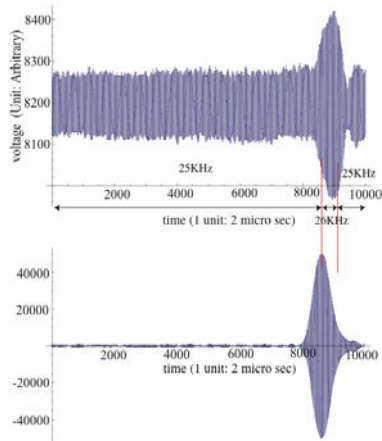


Fig. 2. Received signal and cross-correlation processing result

where r_2 and r_3 are calculated later by measurement of time of flight of ultrasonic waves. The x and y coordinates of transmitter T can be obtained by solving following three equations:

$$\begin{aligned} x^2 + y^2 - r_1^2 &= 0. \\ (x - d)^2 + y^2 - (r_1 + r_2)^2 &= 0. \\ (x + d)^2 + y^2 - (r_1 + r_3)^2 &= 0. \end{aligned} \tag{4}$$

From (4) x and y will be written as:

$$x = \frac{(r_3 - r_2)(r_2 r_3 + d^2)}{2d(r_2 + r_3)}. \tag{5}$$

$$y = \frac{(r_3^2 - d^2)(r_2^2 - d^2)[4d^2 - (r_3 - r_2)^2]}{2d(r_2 + r_3)}. \tag{6}$$

In (5) and (6), r_2 and r_3 can be calculated as:

$$r_2 = V \cdot \Delta t_{R2R1}. \tag{7}$$

$$r_3 = V \cdot \Delta t_{R3R1}. \tag{8}$$

where V indicates the velocity of ultrasonic waves, Δt_{R2R1} corresponds to the time delay of received signal at sensor R2 in respect to sensor R1, and Δt_{R3R1} corresponds to the time delay of received signal at sensor R3 in respect to sensor R1.

3 Experimental Result and Remarks

Resonance type ultrasonic transmitter emits 16 burst pulses at 50 Hz. In burst mode, Sensor dynamics make amplitude of initial rising few waves small compared with successive waves. Small amplitude make small contribution in cross-correlation processing, therefore peak position of cross-correlation which is arrival time of ultrasonic wave, have tendency to shift by few waves. The ultrasonic transmitter pulsates contiguously in order to improve rising property, in burst period by resonance frequency, in idle period by 1 KHz off from resonance frequency.

A 2 GHz PC/AT computer acquires receiver signals via a high speed 1MHz A/D convertor. It then performs signal enhancement on digitized signals and sends the guidance information through FM radio transmission back to the users. Using above mentioned hardware, some test results are shown in Fig. 3 and Fig. 4.

Fig. 3 shows raw position data obtained by the system for nine fixed user locations. No noise cancelling is used in this graph. Positions calculated by the system are shown as cluster of black solid circles.

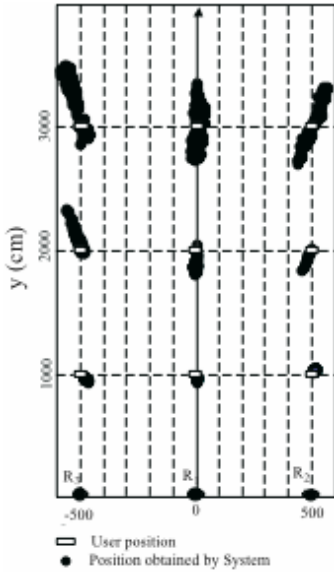


Fig. 3. Stationary test result

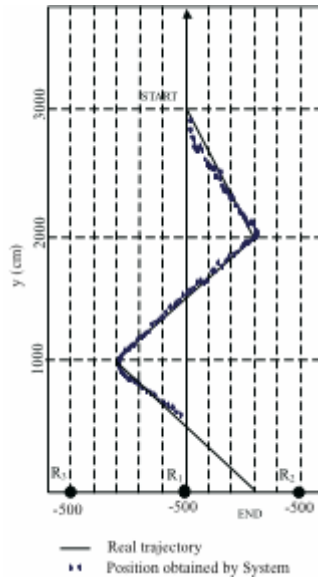


Fig. 4. Motion test result

Fig. 4 demonstrated the results when user is in motion. Here, user starts from the start point and after walking along the zigzag path and reaches to the end point and shows the results after data processing like averaging and noise cancelling. Average is calculated over each 10 incoming data.

In next stage, megaphone will be used for amplifying received signal, though reducing detect angle. Five megaphones (two and three in two line) are installed for making detect angle wide as 75 degree. Each sensor signal installed in each megaphone will be processed by distributed dsPIC controller to detect most reliable signal.

References

1. John, A.B.: New Developments in Mobility and Orientation Aids for the Blind. *IEEE Trans. Biomedical Engineering* 29(4), 285–289 (1982)
2. Aoki, T., Yonezawa, Y., Itoh, K., Hashimoto, M.: Intelligent guide system by parametric sound beam for the visually impaired. In: *Proceedings of the 5th International Conference on Computers Helping People (ICCHP) with special needs*, pp. 637–643 (1996)
3. Petrie, H., Johnson, V.: Evaluation methodologies for navigational aids for blind people. In: *Proceedings of 5th International Conference on Computers Helping People (ICCHP)*, pp. 637–644 (1996)

Pedestrian Navigation System Using Tone Gradient and Robotic GIS

Takafumi Ienaga^{1,2}, Yukinobu Sugimura^{2,3},
Yoshihiko Kimuro^{1,2}, and Chikamune Wada³

¹ Faculty of Information Engineering, Fukuoka Institute of Technology,
3-30-1, Wajiro-higashi, Higashi-ku, Fukuoka, 811-0295, Japan

{ienaga,kimuro}@fit.ac.jp

² Institute of Systems, Information Technologies and Nanotechnologies,
2-1-22, Momochihama, Sawara-ku, Fukuoka, 814-0001, Japan

³ Graduate School of Life Science and Systems Engineering, Kyushu Institute of
Technology, 2-4 Hibikino, Wakamatsu-ku, Kitakyushu, 808-0196, Japan
sugi_kyutech@live.jp, wada@life.kyutech.ac.jp

Abstract. In this study, we propose a new type of pedestrian navigation system for the visually impaired, and which uses “grid maps of accumulated costs” and “tone gradient” techniques. And, we developed the navigation system using a platform called “The RT-middleware”. The developed system was demonstrated at an exhibition of welfare equipments and devices in Japan. And more, we conducted an experiment to evaluate our method for navigation. The results of the experiment suggested that our method was more comprehensible than a method using the oral explanations. And, the results also suggested that the pedestrians felt safer during walk with our method than the oral explanation method.

Keywords: Tone Gradient, R-GIS, Visually Impaired, RFID, RT-middleware.

1 Introduction

Various types of travel aid systems and devices for the visually impaired have been studied and developed. For example, devices which are used as substitution of a white cane and a voice guide system using GPS [1,2], an infrared device [3], a RFID (Radio Frequency IDentification) tag [4], and others [5,6,7] have been developed.

We also studied about auditory maps [8] and devices which indicate direction [9,10] in order to help the visually impaired. The route guidance is part of the travel aid system. The pedestrians are instructed about the route which is selected by the system or themselves. Usually, instructions of the route guidance are constructed by some keywords such as distances, turning points and so on. In this type of instruction, the pedestrians are forced to walk along the selected route, because they have no information except for the selected route. Therefore,

if they lost their way, they are hard to recover to the selected route without any additional information. And, they are required to send requests to the system for finding new route, even if they feel it troublesome. We consider that this fact causes the difficulty of a side trip against a pedestrian who is the visually impaired.

In this study, we propose the new type of pedestrian navigation system for the visually impaired, and which uses “grid maps of accumulated costs” and “tone gradient” techniques. The technique of the grid maps which have accumulated costs is used in robotic technology field. Therefore, to achieve the efficient development, we used a platform called “The RT–middleware” [11]. And, we consider that the navigation using tone gradient technique solves the above problem of recovering to the selected route.

2 System Development

2.1 Basic Plans and System Overview

In order to navigate a pedestrian, the navigation system requires two type of information, that is, the spatial information about a route and the information about the pedestrian’s position and orientation. To generate and convert the former information, we used R-GIS (Robotic and Geospatial Information System) [12]. And, to acquire the latter information, we developed the wireless RFID tag reader system which is able to be attached to a white cane. Because of consideration to the easiness for the functionally expansion, we used the RT-middleware to developed some softwares; R-GIS, wireless tag reader system, and navigation system.

The structure of software components is shown in Fig. 1. Our system has three parts, that is, R-GIS part, a position and orientation measurement part, and a navigation part. We will describe at the following sections.

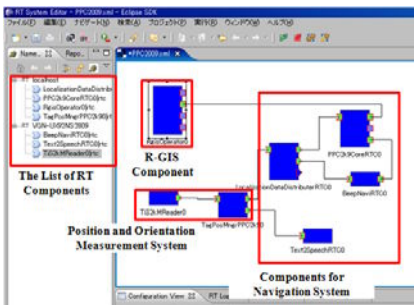


Fig. 1. Software components

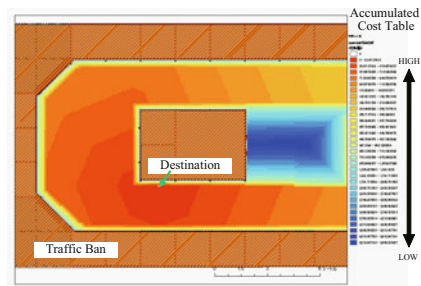


Fig. 2. Example of the accumulated cost

2.2 R-GIS Part

R-GIS provides the functions of various spatial processes and the functions of sharing spatial data to sensors, mobile robots, and other devices.

As one of the providing function, R-GIS generates the grid map which has the accumulated costs from the destination (Fig. 2). The accumulated costs are calculated based on the data in spatial databases and the characteristics of the device. The data in spatial database including costs is able to edit using general GIS (Geographic Information System) software, such as ArcView[13], Quantum GIS[14], and so on. Therefore, we are able to use various GIS data in the world.

And more, a cell size and an extent of the grid map are variable. Therefore, the clients of the R-GIS are able to acquire useful maps whenever they need.

2.3 Position and Orientation Measurement Part

Hardware. We developed wireless RFID tag reader system which is able to be attached to a white cane (Fig.3). The developed system is consist of a RFID tag reader (Texas Instruments: RI-STU-MRD1), a wireless module (Logical Product: LP-RF24-ROBO-01) and other peripheral circuits. This wireless system communicates with the host of wireless module in a serial communication. The host of wireless module is connected to PC (Sony: VGN-UX91NS) by USB cable.

Software. This part is constructed by two RT-Middleware components. One is “TIS2kMReader” and another is “TagPosMngrPPC2k9”. “TIS2kMReader” receives detected tag data and send this data to “TagPosMngrPPC2k9”. In “TagPosMngrPPC2k9” component, an ID of the detected tag is converted to the position data by matching the spatial database. Here, we assume that a pedestrian which has the visually impaired swings a white cane at his/her walking direction. That is, the position which is based on the detected tag shows the next position of the pedestrian. Therefore, the system is able to know the orientation of the pedestrian.

2.4 Navigation Part

The system provides to a pedestrian two types of information. One is the symbolic information about the place name, for example, an entrance of “A” building, in front of the toilet, an entrance of a coffee shop, and so on. The place name is searched by an ID of the detected tag in the spatial database. This function is provided by “TagPosMngrPPC2k9” and “Text2SpeechRTC”.

Another one is the quantitative information about the accumulated cost of the next position by a tone signal. This function is provided by “PPC2kCoreRTC” and “BeepNaviRTC”. The accumulated cost is acquired from the grid map provided by R-GIS. As the next position has lower accumulated cost, the system beeps at higher frequency. As a result, the system is able to generate the tone gradient map. The tone gradient map just looks like the hill climbing algorithm.



Fig. 3. Wireless tag reader system at-

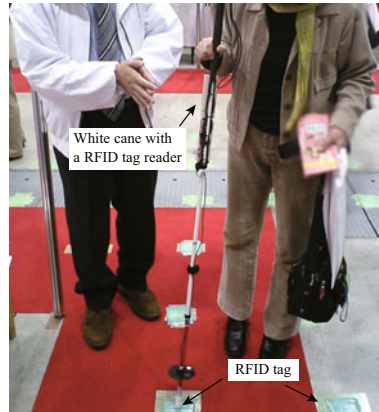


Fig. 4. A snapshot of the demonstra-
tion

Therefore, we are able to expect that he/she reach the destination, even if the pedestrian who is the visually impaired ambles on the way of the destination. This is the most important merit of our navigation system.

And more, a method of our navigation is able to show the most adequate direction of the next step directly. Because, in our system, the most adequate direction of the next step is shown as the highest frequency tone in surrounding the current position. It is expected that our method find his/her walking direction easily than an oral expression using a analog clock dial, an angle and so on.

2.5 Demonstration

We carried out the demonstration at the one of the largest exhibitions of welfare equipments and devices in Japan (PPC2009). This exhibition was held on November 27–29th, 2009. We placed RFID tags (Texas Instruments: RI-TRP-4FF) on the floor for this demonstration. And, the participants were able to experience our system (Fig.4). As the result of the demonstration, we confirmed that the system worked well. And, most of about 5 persons who experienced our system and more than 10 persons who were explained about our system were interested in our system.

3 Experiment

3.1 Purpose

We conducted an experiment to evaluate our navigation method. Here, we evaluated accuracy of direct indication and self-confidence of results.

3.2 Method

We put 100 RFID tags on the floor like a square lattice, and the distance between tags was 0.25[m]. And, the tags were covered by 25 carpets. The arrangement of the tags and a scene of the experiment are shown in Fig. 5.

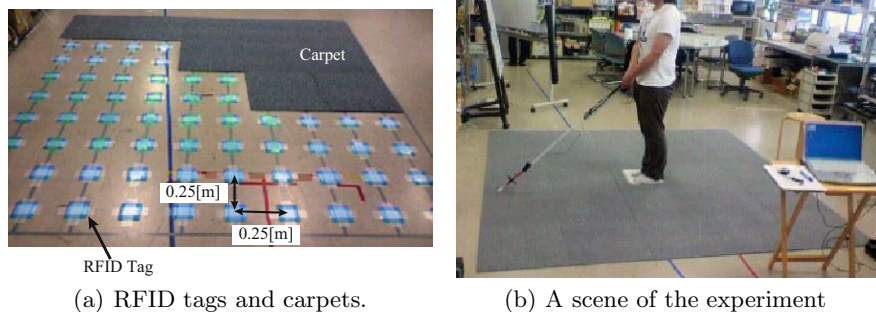


Fig. 5. An arrangement of tags and carpets

We developed the direction indicating system in order to do this experiment. This system used the same method of the above mentioned navigation system to indicate the target direction. That is, as the direction of a subject got closer to the target direction, this system beeped in higher frequency.

We prepared two experimental conditions shown as below.

Oral condition: A target direction was indicated by oral expression, such as three o'clock.

Tone condition: A target direction was indicated by tone signal.

The subjects were five sighted persons. The all subjects did the experiment in "Oral" condition first.

This experiment was carried out in the following procedure.

1. The experimenter explained about our system.
2. A subject was asked to wear a blindfold, and the subject could not get off the blindfold until the experiment was finished.
3. The experimenter brought the subject to an initial position and direction. This initial position was in the center of carpets (Fig.3.2).
4. The subject was asked to change his/her direction to the target direction, after receiving cue from the experimenter.
5. return to number 3.

The target direction was 3, 4, 6, 9 and 10 o'clock. The experimenter selected one of these 5 directions at random. One trial had these five directions. There were five trials for each subject.

We measured a difference between the target direction and the subject's direction. And more, the subject was asked to report his/her self-confidence to his/her decision of the direction in five steps. Here, score "five" meant most confident while "one" meant least confident.

3.3 Results and Discussion

The results of this experiment are shown in Fig.6. A two-way analysis of variance [2(Condition) x 5(Target)] on the total number of errors was conducted. Here, we defined it as an error that the subject turned to the wrong direction against the target direction. "Condition" means the above experimental conditions and "Target" means the above target directions. The main effect of "Condition" was not significant ($F(1,4) = 4.951, p < .1$). And, the main effect of "Target" and the interaction of "Condition" and "Target" were not significant. However, considering the Fig.6(a), it seemed that the subject was able to turn to the target direction correctly in "Tone" condition (3[times]) than "Oral" one(9.8[times]).

A two-way analysis of variance [2(Condition) x 5(Target)] on the confidence score was conducted. The main effect of "Condition" was significant ($F(1,4) = 9.418, p < .05$). And, the main effect of "Target" and the interaction of "Condition" and "Target" were not significant. That is, the subject decided his/her direction more confident in "Tone" condition (4.1) than "Oral" one (3.3) significantly.

As the result of this experiment, it was suggested that our developed system enabled the pedestrians to navigate more comprehensible than the oral expression. And, it was also suggested that the pedestrians were able to walk without worrying than the oral expression.

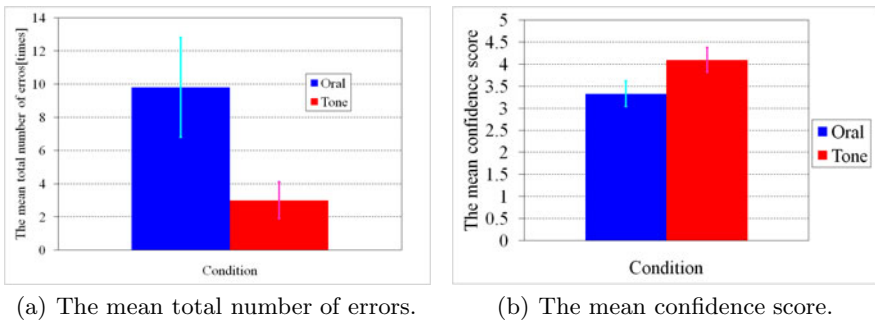


Fig. 6. The results of the experiment

4 Conclusion and Future Works

We proposed the new type of pedestrian navigation system for the visually impaired, which used the tone gradient and R-GIS and showed the effectiveness. In

our future works, we will evaluate our system in the view point of possibility of a side trip. And, we will also improve our system to be able to provide the navigation service in wider area. And more, our system is just a prototype system now. Therefore, the interface of selecting the destination is for sighted persons. The improvement of the interface is also one of our future works.

Acknowledgement

The part of work described in this paper was funded by the Strategic Information and Communications R&D Promotion Programme of the Ministry of Internal Affairs and Communications of Japan and by Intelligent Robot Technology Software Project of New Energy and Industrial Technology Development Organization of Japan. The authors thank R-GIS consortium members, Dr. Arita, Dr. Teraoka, Mr. and Mrs. Araya, for development of R-GIS system. And, we also thank Logical Product Corporation for supporting to develop the wireless RFID tag reader system.

References

1. Adaptive Technology Consulting – A GPS system for the Blind, <http://www.adaptivetech.net/products/trekker.htm> (last access 2006-04-15)
2. Strothotte, T., Petrie, H., Johnson, V., Reichert, L.: MoBIC: user needs and preliminary design for mobility aid for blind and elderly travelers. In: *The European Context for Assistive Technology*, pp. 348–352 (1995)
3. Hatakeyama, T., Hagiwara, S., Koike, H., Itoh, K., Okubo, H., Bond, C.W., Kasuga, M.: Remote Infrared Audible Sign System. *Journal of Human Interface Society* 3(3), 163–170 (2001) (in Japanese)
4. Willis, S., Helal, S.: RFID Information Grid for Blind Navigation and Wayfinding. In: *The Proceedings of the 9th IEEE International Symposium on Wearable Computers (ISWC 2005)*, pp. 34–37 (2005)
5. Borenstein, J., Ulrich, I.: The GuideCane – A Computerized Travel Aid for the Active Guidance of Blind Pedestrians. In: *Proceedings of the IEEE International Conference on Robotics and Automation*, pp. 1283–1288 (1997)
6. Mori, H., Kotani, S.: Robotic travel aid for the blind: HARUNOBU-6. In: *Proceedings of The 2nd European Conference on Disability, Virtual Reality and Associated Technologies*, pp. 193–202 (1998)
7. Graf, B.: Reactive Navigation of an Intelligent Robotic Walking Aid. In: *Proceedings of the 10th IEEE International Workshop on Robot-Human Interactive Communication*, pp. 353–358 (2001)
8. Ienaga, T., Matsumoto, M., Shibata, M., Toyoda, N., Kimura, Y., Gotoh, H., Yasukouchi, T.: A Study and Development of the Auditory Route Map Providing System for the Visually Impaired. In: Miesenberger, K., Klaus, J., Zagler, W.L., Karshmer, A.I., et al. (eds.) *ICCHP 2006*. LNCS, vol. 4061, pp. 1265–1272. Springer, Heidelberg (2006)
9. Asonuma, M., Matsumoto, M., Wada, C.: Study on the Use of Air Stimulation as the Indicator in an Obstacle Avoidance System for the Visually Impaired. In: *The Society of Instrument and Control Engineers Annual Conference 2005*, MA2-14-2 (2005)

10. Ienaga, T., Matsumoto, M., Toyoda, N., Kimura, Y., Gotoh, H., Shibata, M., Yasukouchi, T.: Development of a Direction Indicating System for a Tele-Support for the Visually Impaired. In: ISITA 2008, pp. 612–617 (2008)
11. Ando, N., Suehiro, T., Kitagaki, K., Kotoku, T., Yoon, W.: RT-Middleware: Distributed Component Middleware for RT (Robot Technology). In: International Conference on IROS, pp. 3555–3560 (2005)
12. Kimuro, Y., Araya, R., Arita, D., Ienaga, T., Murakami, K., Yang, Z.: R-GIS: Robotic and Geographic Information System. *Journal of the Robotic Society of Japan* 27(8), 868–876 (2009) (in Japanese)
13. ArcView — Overview, <http://www.esri.com/software/arcgis/arcview/index.html> (last access 2010-1-28)
14. Welcome to the Quantum GIS Project, <http://www.qgis.org/> (last access 2010-1-28)

MOST-NNG: An Accessible GPS Navigation Application Integrated into the MOBILE Slate Talker (MOST) for the Blind

Norbert Márkus¹, András Arató¹, Zoltán Juhász²,
Gábor Bognár, and László Késmárki³

¹ Laboratory of Speech Technology for Rehabilitation,
KFKI Research Institute for Particle and Nuclear Physics, Budapest, Hungary

² Department of Electrical Engineering and Information Systems,
University of Pannonia, Veszprém, Hungary

³ Nav 'N Go Hungary LTD, Budapest, Hungary

Abstract. Over the recent years, GPS navigation has been attracting a growing attention among the visually impaired. This is because assistive technologies can obviously be based on commercially available solutions, as the GPS capable hand-held devices entered the size range of the ordinary mobile phones, and are available at an ever more affordable price, now providing a real choice for a wider audience. For many, an accessible GPS navigator has even become an indispensable tool, an integral part of their every-day life. Since the most appropriate (or at least the most favored) device type for GPS navigation is the so-called PDA, whose user interface is dominated by a touch screen and usually lacks any keyboard, accessibility for the blind remains an issue. This issue has successfully been tackled by the MOST-NNG project in which, the MOBILE Slate Talker's blind-friendly user interface has been combined with Hungary's leading iGO navigator.

1 Introduction

MOST-NNG has been launched in September 2008 as a spin-off of the MOST2 project [2] and aimed at making the widely used map-based iGO GPS Navigator [4] accessible for the blind. The project's objectives were the accessible integration of this GPS navigation program into the MOST framework's non screen reader environment by designing and developing a fully blind-friendly user interface and adding all the necessary extra functionality.

The adequate type of hardware proved to be a small-size off-the-shelf PDA device with an internal GPS receiver, cellular phone and a special mask placed over its touch screen allowing for Braille entry. Such a device powered by the MOST software can function as a talking mobile phone, a memo recorder, a notetaker, etc., and now, as a GPS navigator.

Following the first and most essential software development stage, by early 2009, MOST users with the appropriate PDA device were invited to participate in the testing activity. By March 2010, over 20 blind test users have joined the MOST-NNG

program from all over Hungary. Their feedback plays a vital role in the software development work.

The blind-friendly novel solutions in the domain of accessible ergonomics are the obviously useful byproducts of the project. According to our recent survey among our test users, the simplistic and unified MOST user interface [1] appeared to be favorable for the aged or for users having no or little experience with the computer, as its usage required no extensive training or a demanding learning procedure. The same survey has underlined another benefit of this approach, namely that it obviously encouraged the daily usage of Braille among the users.

2 State of the Art

MOST-NNG based on the MOBILE Slate Talker system is a small cell phone size blind-friendly device unparalleled in its category, and the cheapest solution available today. For the Hungarian blind, MOST is available free of charge. PDAs like Asus P552W and alike are the most adequate hardware with built-in 3G cellular phone and GPS receiver, audio recording and playback capabilities in the price range of \$300-400. MOST is a stand-alone software suite of dozens of useful applications requiring no screen reader. Its simple use and Braille input makes it popular among the elderly. A special Braille mask over the touch screen and a JRE costs the users \$25. Our registered blind testers can use the NNG GPS navigation service free of charge. No wireless data connection is required to access the map data.

Icon Braille+ Mobile Navigator retailed by APH is a compact and versatile PDA-like non-screen reader-style proprietary device for the blind mainly aimed at the English speaking market. It allows for various ways of Braille entering, but the focus is put on wireless connectivity and accessing high quality E-content services, however, the device has no cell phone or GPS receiver. Due to its greater size and weight, Braille+ is not a typical pocket device. Its usage is sophisticated but it requires advanced manual skills and experience. The recent price is \$1,395 plus shipping costs and a mandatory delivery insurance fee. Customers outside the US may also have to pay import/export duties.

A smart phone based solution may consist of a Nokia S80 model, Nuance's Talks mobile screen reader [7] and the Wayfinder Access GPS navigation software [6]. This solution lacks the comfort of Braille entering and due to a screen reader involved, its usage is anything but simple for the blind who are often challenged with the constraint of visualization. The total price of such combined products usually reach or exceed \$1,000-1,200. For the older Symbian phones, an external GPS receiver unit is also required. Wayfinder charges varying fees for map regions periodically and requires a wireless data connection (e.g., GPRS or 3G) to access its on-line maps.

3 How It Works

The choice to integrate the NNG engine into the MOST framework has been a radical one. It immediately meant that no graphical information would be presented to the user. Instead, any pieces of data are channeled via pre-recorded and synthetic speech.

Features challenging the blind users, such as icons, graphical gauges and layout changes, short time-out and type prediction have been replaced with fully accessible, intuitive UI elements that only require the use of the arrow keys and push buttons of the device and Braille holes provided on the touch screen.

Through the program's menus, the user may query different kinds of information related to a geographical location, can look around virtually, exploring the "terrain" either surrounding the current GPS position or any location on the map. So-called monitors may be enabled to keep track of different types of information and to notify the user whenever a value of interest changes. It makes MOST-NNG a helpful long-distance travel companion, even without having to be touched.

When in the public transportation, such features give blind users more confidence for example, to decide exactly when it's time to prepare for getting off the vehicle. So they are no longer forced to exclusively rely on the on-board passenger information system or on the assistance of their fellow passengers.

Besides the classical real-time guidance on planned routes created via MOST's smart search functions, the program offers detailed off-line route itinerary and finely controlled simulation functions, allowing the desired route to be iterated from the comfort of the user's home with the highest possible accuracy.

The GPS navigation plug-in [3] has also been linked to the MOST system's address-book to facilitate the storage and search for geographical locations.

This map-based accessible GPS application has benefited to the users' mobility, also encouraging them to travel to unknown locations autonomously. Currently, it shows the best performance in vehicles in which the user is travelling as a passenger. However, several testers also have satisfying experiences when using it as a pedestrian. There is a growing demand for the enhancement of the pedestrian functionality on behalf of our test users.

4 Special User Interface Solutions

The MOBILE Slate Talker framework [1] provides a user interface that allows the blind and visually impaired users to operate the touch screen dominated PDA devices by simple and natural touch and push actions, and by giving them highly responsive and informative audio feedback. The GPS navigation program is fully integrated into the MOST system, which means that there is practically no need to learn anything new in order to be able to use it.

4.1 Four-Arrow Menu Navigation and Braille Input

In the MOST system, all applications, settings and features are organized into a tree structure and can be accessed via a unified menu system corresponding to this structure. This menu tree can be explored by the sole use of the four arrow keys. The up and down arrows iterate through the items on the same menu level. The right arrow executes the function of the current item and/or enters the underlying submenu. The left arrow in turn leaves the current submenu, going up a level in the menu tree.

Character input is realized by an on-screen Braille entry. A special mask is placed over the PDA's touch screen, turning it into a digital Braille slate. Thereby the user

can only touch the designated areas of the screen through the holes on the mask. During Braille text input, the arrow keys are used for cursoring within the text.

The recent survey shows that both this simplicity of operation and the availability of the Braille entry mode are rewarded by an overwhelming majority of our testers.

4.2 Advanced UI Elements Inspired by the GPS Navigation Interface

A value indicating submenu allows for a sort of simple bi-directional communication. A choice for a certain value or option is typically made from a submenu of two or more elements. When the submenu in question is there to specify a certain setting, the user is informed about the currently set value in such way that when entering the submenu, the focus is immediately brought to the item corresponding to the setting that is currently in effect (e.g., to the vehicle type or route planning algorithm used).

Inclusion of the set value in the menu name is applied to save the user from redundant menu activities. When the set value is directly incorporated into the name of the submenu, the user can learn about the current setting without having to go to a deeper menu level.

A shadow text field for value spelling may come in handy when a certain variable value such as a street name or a postal code in a list appears to be unfamiliar. By selecting the item of interest with the right arrow, the user will be taken to a text field in which, the same exact name can be spelled with the cursor.

Compound values may also appear in menu names in a simplified format. Some compound values (for instance a pair of GPS coordinates) are best represented in a small structure of menu elements. The entity as such manifests itself as a menu node (an entry point to a simple submenu). Both the latitude and longitude values are included in the menu name, rounded to two digits after the decimal point. For the users, when on a quick run-through, this is still easy to digest. After entering the submenu, the user will find separate items assigned to the latitude and longitude components, with their full-precision values appearing in the menu names. Selecting one of these items will take the user to a Braille text field, in which the digits may be cursoried or modified, if that makes sense, or a new value may be entered.

4.3 Value Refreshing Menus and Menu Elements

Certain applications and features raise the need that the user can keep track the constant changes of a set of parameters. In the GPS navigation program, this need has come up most prominently in association with the currently guided route's parameters such as the distance to destination or elapsed travelling time.

The program handles dozens of such rapidly changing data organized into thematic submenus. We have found that it is more than unfortunate to pop the new value "under the user's fingertips" immediately as it changes, especially when several associated values keep changing at once. Instead, we give the user the freedom to choose the moment when an updating should take place.

When an information submenu is entered, the user is presented with a snapshot of the most recent values. Since a user with audio feedback can only check one piece of data a time, a full picture may only be made by iterating through the related items of the submenu. Therefore the related data items are never being updated while the

user is staying on the same menu level. Updating is initiated by selecting any of the elements or leaving and reentering the information submenu. When done so, all the related pieces of data will be updated at once.

4.4 Suppressing Monitor Announcements during Menu Navigation

When for example a street or house number monitor is enabled, the user receives spoken announcements whenever the street name or the current house number changes on the way. Simultaneous monitor announcements are queued so that they don't interrupt each other. Most of the time, this is a very convenient behavior, making the monitors the program's most popular feature among the users. However, well in the early stages of the interface development, special efforts had to be made in order to secure first priority to user interaction.

Thus when any Braille hole is touched or key or button is pressed, the currently spoken announcement is immediately muted and the monitor queue is emptied. Plus, the incoming new monitor announcements are suppressed, while the user is navigating in the menu structure with the arrow keys.

4.5 Tape-Deck-Like Playback Controls for the Route Simulation

A relatively new feature allows the blind user to perform a virtual journey along the planned route. The route simulation may be started, paused, resumed or stopped by tape-deck-like playback controls of a dedicated submenu. The user may easily skip to the next or previous maneuver point along the route. While the simulation is running or paused, the user can "look around" and query different types of information about the current location on the route, as if he or she was actually there on the ground with the GPS device.

4.6 The Structure of the Route Itinerary List

The route itinerary list is a typically off-road, off-line feature that is most benefited in the phase of pre-journey planning and route familiarization. The list is built of menu items, each item representing a section of the route. Using the down and up arrows, the user can get an overview of the route to be iterated.

The names of the list items are used to convey all the necessary route information such as street names, distances to go, types of maneuvers to follow, etc. Each list item constitutes a grammatically correct, easy-to-follow sentence nearing the natural human narration. The most important pieces of information appear at or near the beginning of the sentences. We made it sure that the wording suits both pedestrians and vehicle users. Updating is always initiated by the user, rather than it is done automatically by the program.

5 Users' Feedback

It has been nearly a year since the results of the MOST-NNG project have been available for our Hungarian test users. The test activity takes place on a voluntary basis initiated by the would-be user. In most cases, visually impaired individuals join the

test activity after they got in touch with the MOST2 team. They carry their own PDA devices to the support center where the MOST-NNG software and the Braille mask will be installed, then the users receive a two hour long personal training that enables them to go on with using the MOST-NNG device autonomously.

An initial data inquiry on the users' background takes place during this first personal encounter. Then it is followed by an active communication with many of the testers via E-mail, telephone or in person.

For us, this year has been a great experience of a kind of collective thinking. Any new feature proposed by the testers is added with special emphasis on its ergonomical relevance, and with the hope that it will facilitate the users' mobility. However, it often happens that our testers articulate a demand whose realization is already on the way. That gives us an assurance that we are going to the right direction.

At the time of the writing of this paper, we have knowledge about 21 MOST-NNG test users living in 8 counties of Hungary. 48% of them live in the capital Budapest, a further 38% in 4 central and Mid-Western counties, while the remaining 14% are strewn over the North-Eastern, Eastern and Southern regions of the country. The majority of the users are concentrated at the economically most developed urban areas of the country, as 48% live in the largest city, 19% in other major cities, 24% in smaller towns and only 9% in townships/villages.

Another important finding is the testers' age distribution ranging from 17 to 74 years and with a mean age of 45 years. The highest peak (29%) is observed at the age group of 25-34 years. The next three age groups (35-64 years) represent 14.3% each, roughly making 43% of the testers. There is another peak of 19% at the group of 65-74 years, and only 9% belongs to the youngest measured age group of 15-24 years.

The two peaks (at around 30 and 70) correspond to the main target groups. The youngest generation appears to be underrepresented, as they tend not to learn and use Braille, and fewer of them travel autonomously (i.e., without sighted assistance).

Our previously existing MOST users (mainly in the age range of 45-65 years) make 29% of the recent MOST-NNG testers, whereas the remaining 71% consists of new comers. 33% of the testers (mostly over the age of 55) had previously no or very little experience with the computer when they decided to join this activity. Even for them, the two hour preliminary training has turned out to be sufficient.

24% were not familiar or had not used Braille before joining this program. Although the GPS navigation functions could also be used without this skill, in a matter of days or weeks, all of these individuals had learnt and practiced to enter Braille signs on the mask over the touch screen [2].

An important conclusion of this last year is that the users long for constancy. Once they got familiar with the program's layout, they become highly sensitive of any changes affecting this "friendly environment". Thus when dealing with blind users, the scope and frequency of general overhauls should be minimized for the sake of the best user experience.

We have also realized that a significant portion of the users require 1-3 months to get to using and exploiting the newly added features. This may correlate with the fact that some of these users travel only once or twice a month with the MOST-NNG device and therefore they don't have so many opportunities for testing.

5.1 Questionnaire Based Survey

In the first two weeks of March 2010, a survey was conducted among the MOST-NNG testers based on a simple questionnaire. We have received valuable feedback from 17 persons (81% of our registered testers). The main goal of the survey was to gauge the testers' most recent opinion specifically about the accessible user interface. Probably due to the high degree of integration, the respondents viewed the MOST-NNG device as a whole and they didn't treat the MOBILE SLATE TALKER [1], the NNG engine or the hosting PDA device as separate products.

From our point of view, it is quite a positive sign that only a small fraction of the criticisms touched on the actual user interface, while the majority raised hardware related issues and those concerning the navigation engine.

As a first task, the respondents have ranked the main factors that have influenced their decision in choosing the MOBILE SLATE TALKER. 72% indicated the Braille entering modality as a first-priority governing factor. 62% ranked the "easy-to-learn usage" as their top priority. 34% thought that there was no other comparable product available on the market. Several users have chosen this device explicitly for its GPS capabilities, while other respondents have praised MOST for its versatility of applications offered in a single device.

18% use the GPS navigation daily, 47% weekly and 35% once or twice in a month. 71% has tried it in pedestrian mode, 58% of them use it rarely while walking, 25% often and 17% do it mostly so. 67% of the testers have experienced GPS accuracy problems at walking speed. Another 25% stated that they were able to compensate the inaccuracies with more attention and experience.

The significance of this issue is that each tester has received the software along with the recommendation to most preferably use it on board of vehicles, since the navigation engine had previously been optimized for such a scenario, but most users actually felt free to test the system under all kinds of travelling circumstances.

On the other hand, all respondents used the MOST-NNG system in vehicles, 29% exclusively so, 35% mainly, 24% often and 12% just rarely. An overwhelming majority, 82% used it on buses (71% on local buses, 41% also on long-distance buses), 53% in other kinds of local public transportation, 47% in trains and 47% in cars.

88% stated that the GPS navigation program mostly helped them in determining their current location. 59% specified that the program was most used in the public transportation to keep track of the stops or stations and to find out when to get off. 35% simply highlighted its facilitating role in orientation, while 24% emphasized its use in pre-journey route preparations.

71% of the respondents gained experience with route planning and its related functions (guidance, simulation, itinerary, etc). 58% of them preferred the direct search for postal addresses as the main method for building routes, 50% favored searching among the locations previously saved in MOST's integrated address book, and 22% mostly obtained the targets for their routes by searching for POIs (Point Of Interests). Cultural institutions, cash-points, administrative offices and shops were among the most popular points of interest mentioned by the respondents.

52% found the route itinerary list, and 41% the route simulation useful tools. Actually, the rest of the testers had not been able to try these relatively new features out by the time of the survey due to various reasons.

Concluding the questionnaire, the testers were asked about what else they would like to have in the GPS navigation program on top of the existing features and capabilities. 59% expressed their demands for a greater GPS accuracy at walking speed. 41% declared that they were satisfied with MOST-NNG's current performance. A single user asked for a simpler menu structure, while the rest of the users have declared to get along well with the user interface of the MOBILE SLATE TALKER.

6 Conclusion

This paper presents a detailed description on the novel solutions implemented to provide full accessibility for the blind to a highly graphical GPS application on a congenitally visual hand-held device, and on how the users' experiences and proposals have been incorporated into the software product.

References

1. Arató, A., Juhasz, Z., Blenkhorn, P., Evans, D.G., Evreinov, G.E.: Java-Powered Braille Slate Talker. In: Miesenberger, K., Klaus, J., Zagler, W.L., Burger, D. (eds.) ICCHP 2004. LNCS, vol. 3118, pp. 506–513. Springer, Heidelberg (2004)
2. Juhasz, Z., Arato, A., Bognar, G., Buday, L., Eberhardt, G., Markus, N., Mogor, E., Nagy, Z., Vaspori, T.: Usability Evaluation of the MOST Mobile Assistant (SlatTalker). In: Miesenberger, K., Klaus, J., Zagler, W.L., Karshmer, A.I. (eds.) ICCHP 2006. LNCS, vol. 4061, pp. 1055–1062. Springer, Heidelberg (2006)
3. Markus, N., Juhasz, Z., Bognar, G., Arato, A.: How Can Java Be Made Blind-Friendly? In: Miesenberger, K., Klaus, J., Zagler, W.L., Karshmer, A.I. (eds.) ICCHP 2008. LNCS, vol. 5105, pp. 526–533. Springer, Heidelberg (2008)
4. Nav N Go KFT, <http://www.navngo.com>
5. Icon Braille+ Mobile Manager, personal digital assistant, http://www.aph.org/tech/pda_doc.htm
6. Wayfinder Access, mobile navigation software, <http://www.wayfinder.com>, <http://www.talknav.com/access.php>
7. Nuance Talks mobile screen reader and Vocalizer TTS engines, <http://www.nuance.com/talks/>, <http://www.talknav.com/talkspremium.php>

Improving Computer Vision-Based Indoor Wayfinding for Blind Persons with Context Information

YingLi Tian¹, Chucai Yi¹, and Aries Arditì²

¹ Electrical Engineering Department
The City College and Graduate Center
City University of New York, New York, NY 10031
ytian@ccny.cuny.edu, cyi@gc.cuny.edu

² Arlene R Gordon Research Institute
Lighthouse International
111 East 59th Street, New York, NY 10022
aarditi@lighthouse.org

Abstract. There are more than 161 million visually impaired people in the world today, of which 37 million are blind. Camera-based computer vision systems have the potential to assist blind persons to independently access unfamiliar buildings. Signs with text play a very important role in identification of bathrooms, exits, office doors, and elevators. In this paper, we present an effective and robust method of text extraction and recognition to improve computer vision-based indoor wayfinding. First, we extract regions containing text information from indoor signage with multiple colors and complex background and then identify text characters in the extracted regions by using the features of size, aspect ratio and nested edge boundaries. Based on the consistence of distances between two neighboring characters in a text string, the identified text characters have been normalized before they are recognized by using off-the-shelf optical character recognition (OCR) software products and output as speech for blind users.

Keywords: Indoor navigation and wayfinding, indoor, computer vision, text extraction, optical character recognition (OCR).

1 Introduction

Based on the 2002 world population, there are more than 161 million visually impaired people in the world today, of which 37 million are blind [1]. Challenges associated with independent mobility are well known to reduce quality of life and compromise the safety of individuals with severe vision impairment. Robust and efficient indoor object detection from images/video captured by a wearable camera has the potential to help people with severe vision impairment to independently access unfamiliar indoor environments. But indoor object detection poses several challenges: 1) there are large intra-class variations of the object model among different indoor environments. The appearance and design style of different instances of objects (e.g. doors) are quite variable in different buildings; 2) there are small

inter-class variations for certain object models. As shown in Figure 1, the basic door shapes of a bathroom, an exit, a lab door, and an elevator are very similar. It is very difficult for a machine-based system to distinguish them without using context information on/around the object; 3) in contrast to objects with enriched texture and color in natural or outdoor scenes, most indoor objects are man-made with less texture. Existing feature descriptors may not effectively represent the indoor objects; 4) it is unrealistic to expect an intact object, especially in its entirety and in a canonical orientation, to be captured by a blind user with a wearable camera that has no reliable means for steering or aiming the camera. Indoor object detection methods for our application must handle situations in which only part of the object is captured; and 5) when the user moves while wearing a camera, changes of position and distance between the user and the object will cause big view and scale variations of the objects.

Signs with text and icons play an important role in guiding us through unknown environments and can be very helpful to persons who are blind, if the signage can be recognized and decoded by a wearable computer vision system. To improve the ability of people who are blind or have significant visual impairments to access, understand, and explore unfamiliar indoor environments, it is essential to incorporate the context information with object detection from multiple visual cues captured using wearable cameras. In this paper, we present a robust and efficient algorithm for text extraction from indoor signage. The extracted text is recognized by the off-the-shelf optical character recognition (OCR) software products and then converted to speech for blind users by existing speech-synthesis software.

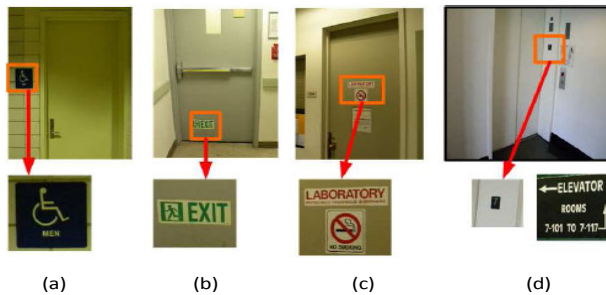


Fig. 1. Indoor objects (top row) and their associated contextual information (bottom row): (a) a bathroom, (b) an exit, (c) a lab room, (d) an elevator. Contextual information (bottom row) must be employed to distinguish them.

2 State-of-the-Art

Camera-based document analysis and recognition have recently gained more attention due to the pervasive use of camera phones and hand-held digital still and video cameras [2-5, 7-8, 14]. Compared to high-resolution and high-quality scanned document images, camera-captured document images will result in poor performance due to: (1) Insufficient resolution, (2) uneven lighting, (3) perspective and camera lens distortion, (4) text on non-planar surfaces, (5) complex backgrounds, (6) defocus (8) camera or document movement and image smear, (9) intensity and color quantization, and (10)

sensor noise. Liang et al. described these challenges with more details in a well-known survey paper [4]. Many methods have been proposed for text extraction, enhancement, and binarization. A survey conducted by Trier and Taxt compared 11 locally adaptive thresholding techniques [10]. They concluded that Niblack’s method [6] is the most effective for general images. However, the method produces a noisy output with homogeneous regions larger than the size of the window. Some improved methods [9, 12] have been proposed by assuming black text on a white background or by measuring local contrast to estimate threshold. However, these conventional methods cannot handle text on cluttered backgrounds. Recently, a number of papers have addressed text extraction from colorful documents [3, 7, 15]. Kasar et al. [3] proposed a straightforward method to robustly extract the text from cluttered backgrounds and convert the text in black on a white background. Nikolaou et al. [7] developed a meanshift method to reduce the number of complex background colors for document analysis.

3 Context Information Extraction from Signage

Signage and other visual information provide important guidance in finding a route through a building by showing the direction and location of the desired end-point. Visual information is particularly important for distinguishing similarly shaped objects such as elevators, bathrooms, exits, and normal doors. Figure 2 shows some examples of signage of indoor environments. In this paper, we focus on text extraction and recognition from signage. Our method includes three components: 1) color reduction; 2) text extraction and normalization; and 3) text recognition.

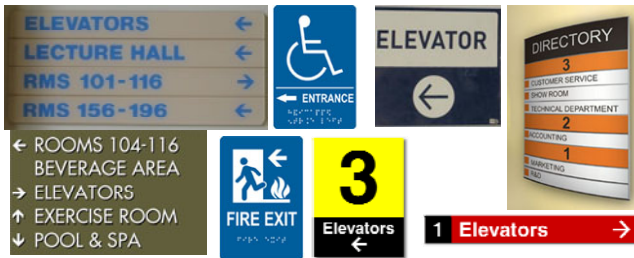


Fig. 2. Examples of signage in indoor environments

3.1 Color Reduction for Text Extraction

To segment and extract text from images containing multiple colors, inspired by [7], we develop a color reduction method as a quantization preprocessing stage based on edge and color information. First, an edge preserving smoothing filter is applied to the entire image to remove image noises without deforming the details of object boundaries. Second, in order to represent the colors of all the objects on the image, the RGB color distribution of the image is sampled by selecting only those pixels which are local minima in the 8-neighborhood on the edge map image. This process ensures that the samples are non-edge points, which guarantees that they are not on the boundary

of the objects. Thus, fuzzy points on the transition areas between objects on the image are avoided. Third, the sampled pixels with similar colors are grouped into color clusters. Fourth, a mean-shift operation is performed on the obtained color clusters to further merge similar color regions in RGB space and produce the final number of colors. This final number of colors should be very small (< 10). At this point, the processed image should have solid characters and uniform local backgrounds for text extraction.

After color quantization, with the purpose of classifying the information with different colors, each quantized value is placed against a white background, as shown in Figure 3. Obviously color reduction decomposes image into several color layers. Each layer contains exactly two colors in the form of informative foreground and white background. The interference from background is reduced for text extraction.



Fig. 3. Examples of color reduction. The 1st column shows the original images; the 2nd and 3rd columns show color layers after color quantization.

3.2 Text Extraction

To extract the regions containing text from cluttered backgrounds, we apply a text extraction method in each color layer which is obtained from color reduction processing. Generally, text regions have more transitions between foreground (e.g. text) and background than non-text regions. Based on this observation, the maximum gradient difference (MGD) [13] for each pixel is calculated to describe the density of binary switches, which are transitions from foreground to background and vice versa in each color layer of the input image, where the foreground is denoted by 1 and the background is denoted by 0. A 1×9 sliding window is defined to detect the transitions around each pixel, as shown in Figure 4(a). The MGD of each pixel is calculated by Equation (1).

$$MGD(P_c) = \begin{cases} 0, & \sum_{k=-4}^4 P_{c+k} = 0 \text{ or } \prod_{k=-4}^4 P_{c+k} = 1 \\ 1, & \text{otherwise} \end{cases} \quad (1)$$

where P_c is a pixel located in the center of the sliding window $\{P_{c+k} | -4 \leq k \leq 4\}$.

In the MGD map, regions containing text information correspond to clusters of value 1. Therefore we search the MGD map for the rows and columns which compose text regions by summing the MGD values row-by-row and column-by-column, as presented by Equation (2).

$$R_H(i) = \sum_j MGD(P_c(i, j)) , R_V(j) = \sum_{R_H(i) > T_H} MGD(P_c(i, j)) \quad (2)$$

where $R_H(i)$ is the accumulation of MGD values in the i th row and $R_V(j)$ is that in the j th column. If $R_H(i)$ and $R_V(j)$ are greater than predefined thresholds T_H and T_V respectively, the corresponding rows and columns are components of text regions, as shown in Figure 4(b).

In order to detect more complex text with irregular size and shape, we extend the algorithm by using a text structure model to identify letters and numbers. This model is based on the fact that each of these characters will contain no more than two closed edge boundaries (e.g., “A” contains one such boundary and “8” contains two). First, the edge map of the text regions is computed, as shown in Figure 5(c-d). The edge bounding box is obtained by performing connected component labeling. A simple ratio filter is then applied to filter out the obvious non-text regions, as shown in Figure 5(e-f). To convert the text to black on a white background, Kasar *et al.* [3] employed an edge-based connected component approach to automatically determine a threshold for each component of RGB channels. In our experiment, each color layer has already been a binary image, so we only set the foreground pixels as black and background as white.

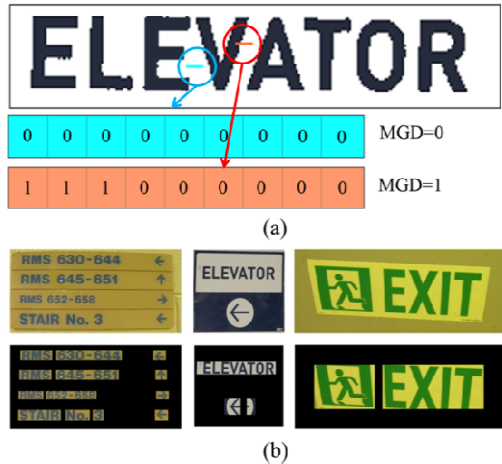


Fig. 4. (a) Examples of sliding windows in a text region; (b) Text regions detected by MGD algorithm. Top row is original image; bottom row shows identified text regions against a black background.

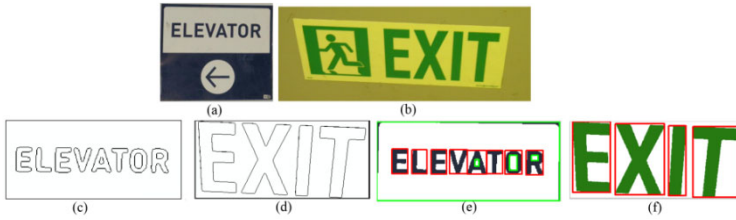


Fig. 5. (a-b) Original images; (c-d) Edge map of some text regions; (e-f) Edge bounding boxes where the red boxes indicate the regions of text characters. The green boxes are the non-text regions which filtered out by predefined rules.

3.3 Text Normalization and Recognition

Off-the-shelf optical character recognition (OCR) software for text recognition has constraints with respect to font size and orientation variability. Before we input the text image to OCR, we must normalize the text to the size range which these products can handle. We use more text structure features based on the geometric relationships among three neighboring characters to further filter out the non-text bounding boxes and normalize text characters. As shown in Figure 6, the geometric relationship includes the consistency of distances between centroids. The normalized text strings are then recognized by OCR and displayed as speech to blind users.

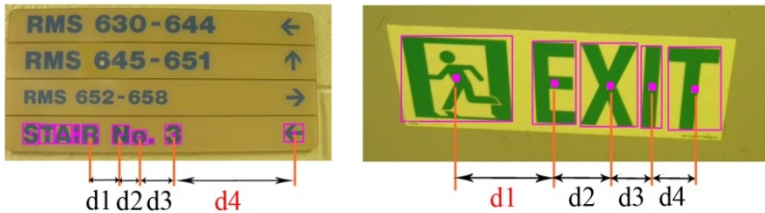


Fig. 6. Text bounding boxes share consistent distances while non-text bounding boxes are connected by inconsistent distance denoted by red

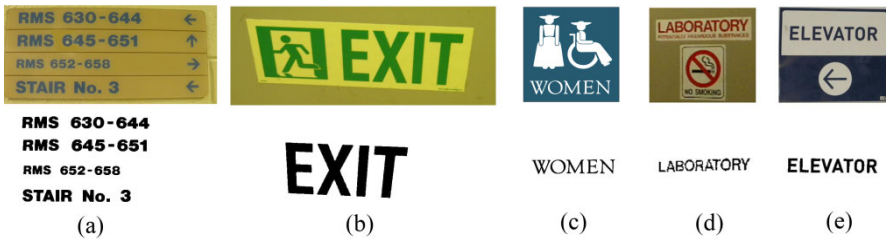
4 Experiment Results

Our method is robust and efficient for text extraction and recognition and is evaluated on a database of signs in indoor environments. Figure 7 shows some examples of the results for text extraction and recognition from indoor signage. The extracted text characters are recognized by using an off-the-shelf OCR software product. In our experiment, OmniPage Professional Version 16 is used to transform the extracted text from image regions to readable ASCII codes.

Table 1 demonstrates the results of OCR detection corresponding to Figure 7. The context information, including both letters and digits, is accurately extracted by our method. Note that our text extraction makes using standard OCR feasible. Images of signs from a camera without this process would surely fail or provide inaccurate character recognitions.

Table 1. Results of text recognition of images in Figure 7.

Text images	OCR outputs
Figure 7(a)	RMS 630-644 RMS 645-651 RMS 652-658 STAIR NO.3
Figure 7(b)	EXIT
Figure 7(c)	WOMEN
Figure 7(d)	LABORATORY
Figure 7(e)	ELEVATOR

**Fig. 7.** Examples results of text extraction for signage in indoor environments

5 Impact and Contributions

We have developed a new framework for text extraction and recognition. Context information, including that provided by indoor signage, plays a very important role for navigation and wayfinding of blind or visually impaired people independently accessing unfamiliar indoor environments. This research has following impact: (1) It significantly enriches the study of object detection by incorporating context information, and leads to significant improvements over existing methods; (2) The method developed in this paper provides new strategies and the promise of new technologies for blind and visually impaired persons to access unfamiliar indoor environments; and (3) The research will benefit many other important research areas including video surveillance, intelligent conference rooms, video indexing, human-computer interactions, *etc.*

6 Conclusion and Future Work

We have described a new method to extract and recognize text from indoor signage. First, we propose a color reduction process to reduce complex colors of both text and background. Then, we apply a robust and efficient text extraction to generate black text on white background. Finally, to achieve high recognition by using existing OCR software, we apply a post-processing step to model geometric relationship among neighbor characters to further filter out the non-text bounding boxes and use the consistency of distances between centroids of these text characters. Our future work will focus on recognizing iconic signage, which of course does not contain text

information. We will also address the significant human interface issues associated with prioritizing which objects and signs are most relevant to the blind user's navigation goals, and with the significant issues relating to auditory display of the recognized information.

Acknowledgments. This work was supported by NSF grant IIS-0957016 and NIH grant EY017583.

References

1. Resnikoff, S., Pascolini, D., Etya'ale, D., Kocur, I., Pararajasegaram, R., Pokharel, G.P., et al.: Global data on visual impairment in the year 2002. *Bulletin of the World Health Organization* 82, 844–851 (2004)
2. Fu, H., Liu, X., Jia, Y.: Gaussian Mixture Modeling of Neighbor Characters for Multilingual Text Extraction in Images. In: *IEEE International Conference on Image Processing, ICIP (2006)*
3. Kasar, T., Kumar, J., Ramakrishnan, A.G.: Font and Background Color Independent Text Binarization. In: *2nd International Workshop on Camera-Based Document Analysis and Recognition (2007)*
4. Liang, J., Doermann, D., Li, H.: Camera-based analysis of text and documents: a survey. *IJDAR* 7(2-3) (July 2005)
5. Lyu, M., Song, J., Cai, M.: A Comprehensive method for multilingual video text detection, localization, and extraction. *IEEE transactions on circuits and systems for video technology* 15 (2005)
6. Niblack, W.: *An introduction to digital image processing*, pp. 115–116. Prentice Hall, Englewood Cliffs (1986)
7. Nikolaou, N., Papamarkos, N.: Color Reduction for Complex Document Images. *International Journal of Imaging Systems and Technology* 19, 14–26 (2009)
8. Phan, T., Shivakumara, P., Lim Tan, C.: A Laplacian Method for Video Text Detection. In: *The 10th International Conference on Document Analysis and Recognition*, pp. 66–70 (2009)
9. Sauvola, J., Pietikainen, M.: Adaptive document image binarization. *Pattern Recognition* 33, 225–236 (2000)
10. Trier, O., Taxt, T.: Evaluation of binarization methods for document images. *IEEE Trans. Pattern Analysis and Machine Intelligence* 17(3), 312–315 (1995)
11. Tran, H., Lux, A., Nguyen, H., Boucher, A.: A novel approach for text detection in images using structural features. In: *The 3rd International Conference on Advances in Pattern Recognition*, pp. 627–635 (2005)
12. Wolf, C., Jolion, J., Chassaing, F.: Text localization, enhancement and binarization in multimedia documents. In: *Proc. ICPR*, vol. 4, pp. 1037–1040 (2002)
13. Wong, E., Chen, M.: A new robust algorithm for video text extraction *PR36(6)* (June 2003)
14. Zhang, J., Kasturi, R.: Extraction of Text Objects in Video Documents: Recent Progress. In: *The 8th IAPR International Workshop on Document Analysis Systems (2008)*
15. Zhong, Y., Karu, K., Jain, A.K.: Locating text in complex color images. In: *Proc. ICDAR (1995)*

Computer Vision-Based Door Detection for Accessibility of Unfamiliar Environments to Blind Persons

Yingli Tian¹, Xiaodong Yang¹, and Aries Ardit²

¹ Department of Electrical Engineering
The City College, City University of New York,
160 Convent Ave., New York, NY 10031
{ytian, xyang02}@ccny.cuny.edu

² Arlene R Gordon Research Institute
Lighthouse International
111 East 59th Street, New York, NY 10022
aarditi@lighthouse.org

Abstract. Doors are significant landmarks for indoor wayfinding and navigation to assist blind people accessing unfamiliar environments. Most camera-based door detection algorithms are limited to familiar environments where doors demonstrate known and similar appearance features. In this paper, we present a robust image-based door detection algorithm based on doors' general and stable features (edges and corners) instead of appearance features (color, texture, etc). A generic geometric door model is built to detect doors by combining edges and corners. Furthermore, additional geometric information is employed to distinguish doors from other objects with similar size and shape (e.g. bookshelf, cabinet, etc). The robustness and generalizability of the proposed detection algorithm are evaluated against a challenging database of doors collected from a variety of environments over a wide range of colors, textures, occlusions, illuminations, scale, and views.

1 Introduction

Independent travel is well known to present significant challenges for individuals with severe vision impairment, thereby reducing quality of life and compromising safety. There have been many efforts to study blind navigation and wayfinding with the ultimate goal of developing useful travel aids for blind people [5, 12], but very few have met with more than limited success. The most useful and accepted independent travel aids remain the Hoover white cane and the guide dog, both of which have been in use for many years. While GPS-guided electronic wayfinding aids show much promise in outdoor environments, there is still a lack of orientation and navigation aids to help people with severe vision impairment to independently find doors, rooms, elevators, stairs, bathrooms, and other building amenities in unfamiliar indoor environments.

In this paper, we develop a computer vision-based door detection method for assisting blind persons to access unfamiliar indoor environments. This method perhaps would be one module within a more complete computer-vision aid designed for blind persons that would mainly consist of a single camera and a computer. For example,

visual information would be captured via a miniature camera mounted on the head via a cap or sunglasses (Figure 1), while image processing and speech output would be provided by a computer (with speech output via headphones or mini-speakers and updated in real-time). Our detection results from a large dataset of indoor images including door images and non-door images demonstrate that the proposed method is robust and generic to the changes of scales, view points, and occlusions in different buildings.

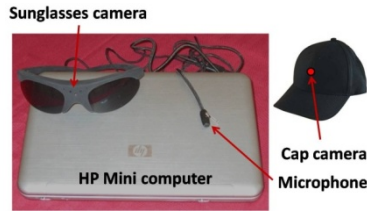


Fig. 1. Computer vision-based indoor navigation and wayfinding prototype system

2 State-of-the-Art

Door detection approaches have been developed for robot navigation. Some of these methods use laser range finders to establish range sensor models of the surrounding environment and to obtain the distance data to test door concavity [1, 7]. Stoeter *et al.* [13] employed sonar data to confirm or dismiss detection results from cameras. In [9], three cameras are employed to perform stereo vision for door detection. However, high-cost, high-power, and complexity in these systems make them inappropriate to work for visually impaired people. To reduce the cost and complexity of the device and its computational requirements, we use a single camera.

There are a few existing door detection algorithms using monocular visual information [3, 10, 11]. In [3] an AdaBoost classifier is trained to detect doors of similar appearance by combining the features of pairs of vertical lines, concavity, gap between the door and floor, color, texture, kick plate, and vanishing point. However, a perceptible gap below the floor and kick plate is not always present in different environments. Munoz-Salinas *et al.* [10] developed a doorframe model-based door detector by using Hough Transform to extract the edge segments and a fuzzy system to analyze the relationship between the segments. However, their algorithm cannot differentiate doors from large rectangular objects typically in indoor environments, such as bookshelves, cabinets, and cupboards. In [11], two classifiers were trained by using color and shape features. This algorithm was designed to detect the doors of the authors' office building, where all the doors have similar color. It would fail if the colors of the doors vary.

To overcome the limitations described above, we develop an image-based door detection algorithm by establishing a general geometric door model that utilizes the general and stable features of doors (i.e. edges and corners) without a training

process. Furthermore, integrated with geometric information of lateral at similar horizontal coordinate, the proposed algorithm is able to distinguish doors from other objects with door-like shape and size. The detection results demonstrate that our door detection method is generic and robust to different environments with variations of color, texture, occlusions, illumination, scales, and viewpoints.

3 Methodology for Door Detection

3.1 Geometric Door Model

The ideal geometric model of a door consists of four 90° corners and four lines of the doorframe. However, due to perspective and occlusion, only part of a door is often captured by a wearable camera, and the corners may depart significantly from 90°. Figure 2 illustrates ideal and more realistic geometric door models including those under ideal conditions (a), and conditions with occlusion and perspective effects (b and c). Each geometric model includes four corners (red square) and four lines (red lines).

Our algorithm makes weak assumptions about doors in an image: 1) At least two corners of each doorframe are visible; 2) both vertical lines of each doorframe are visible; 3) vertical lines of doorframes are nearly perpendicular to the horizontal axis of the image; 4) doors in the image have at least a certain width and length. These requirements are very likely to met in most wearable camera images when a door is present in the field of view, and help make our door detection method more robust to variations of color, texture, occlusion, and door status (open or close). For example, the color of a door can be similar or different from that of a wall. The surface of a door can be with or without texture. The material of a door can be wood, metal, or glass. The status of a door can be closed or fully opened. The upper or lower part of a door can be occluded.

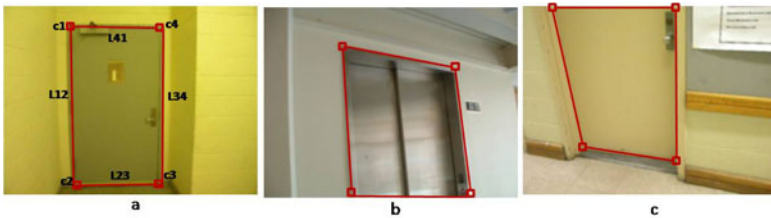


Fig. 2. The proposed geometric door model: (a) The ideal condition without occlusion or perspective effects. (b) The condition with lower part occluded and perspective effects. (c) The condition with upper part occluded and perspective effects.

3.2 Edges and Corners Detection

Edges and corners in images are relatively stable features to identify and are resistant to variations of scales, colors, viewpoints, and light changes. In order to develop a robust door detector to handle different environments, we combine both edge and corner features to characterize the geometric door shape model. In our system, we

first apply Gaussian smoothing to reduce image noise. The smoothing process also eliminates the unnecessary corners for door detection. Then we extract edges through Canny edge detection [2] and create a binary edge map. Then corners are extracted through the edge map based on global and local curvature properties by the corner detector proposed by He and Yung [6]. Since the endpoint of an open contour is also considered as a corner by this method, four corners of a doorframe can be extracted by this corner detector regardless of occlusion (Figure 2(b, c)).

$$Siz_{ij} = \frac{\sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}}{DI} \quad (1)$$

$$Dir_{ij} = \tan^{-1} \frac{|x_i - x_j|}{|y_i - y_j|} \times \frac{180}{\pi} \quad (2)$$

As shown in Figure 2(a), C_1 , C_2 , C_3 , and C_4 are four corners with coordinate of (x_i, y_i) . L_{12} is the line connecting C_1 and C_2 , similarly for L_{23} , L_{34} , and L_{41} . Suppose the length of the diagonal of an image is DI . The ratio between the length of L_{ij} and DI is measured by the variable Siz_{ij} . The direction of L_{ij} corresponding to the horizontal axis of an image is measured by the variable Dir_{ij} . Siz_{ij} and Dir_{ij} are defined in equation (1) and (2), respectively. Then, we use the geometric relationship of any four corners to establish door-corner candidates which satisfy the following rules:

1. A door in an image at least has a certain width and height. So, Siz_{12} and Siz_{34} should be within a certain range:

$$\begin{aligned} HeightThresL < Siz_{12}, Siz_{34} < HeightThresH \\ WidthThresL < Siz_{23}, Siz_{41} < WidthThresH \end{aligned}$$

2. Due to perspective deformation, L_{23} and L_{41} could form a certain angle with the horizontal axis. But, Dir_{23} and Dir_{41} should not be too large:

$$Dir_{23}, Dir_{41} < DirectionThresL$$

3. Vertical lines of a door frame are almost perpendicular to the horizontal axis of the image. So, Dir_{12} and Dir_{34} should be large enough:

$$Dir_{12}, Dir_{34} > DirectionThresH$$

4. Vertical lines of a door frame should parallel with each other:

$$|Dir_{12} - Dir_{34}| < ParallelThres$$

5. The ratio between height and width of a door frame should be within a range:

$$HWThresL < (Siz_{12} + Siz_{34}) / (Siz_{23} + Siz_{41}) < HWThresH$$

3.3 Door Model Matching by Combination of Edges and Corners

Door-corner candidates represent possible geometric models. The next step is to determine whether a door-corner candidate matches to a real door frame by combining the edge information. Most existing door systems employ Hough transforms [8] to detect straight lines. However, practical experiments demonstrate several disadvantages of this approach, such as missing start and end points, sensitivity to parameters, and unwanted merging or splitting of lines. A novel aspect of our algorithm is its use of the edge image as the reference map to match with the door-corner candidates, rather than directly detecting the lines. The proposed matching process can provide

the information of start and end points. Besides, it is insensitive to the broken segments caused by small gaps. The concept of “fill-ratio” is the foundation for the matching process. In Figure 3(a), C_i and C_j are two corners. The gray area is a region formed by masks expanding from the straight line connecting C_i and C_j . In our algorithm implementation, the mask is a 7×7 window. The black line is the detected edge. The edge pixels in the gray region constitute the overlapping line. The “fill-ratio” of corner C_i and C_j in the edge map is defined as:

$$FR_{ij} = \frac{OverLap_{ij}}{Length_{ij}} \tag{3}$$

where $Length_{ij}$ is the length of L_{ij} , the straight line connecting corner C_i and C_j . $OverLap_{ij}$ is the length of the overlapping line.

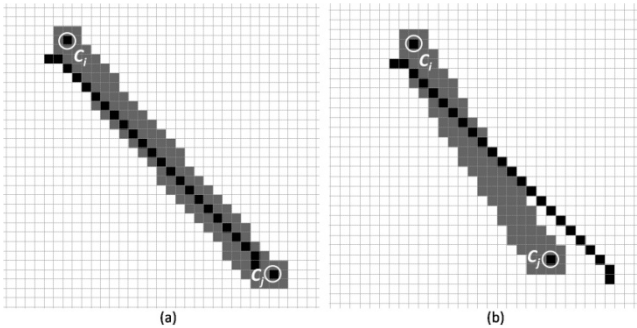


Fig. 3. Definition of “fill-ratio” for matching edges with two corners of a door-corner candidate. C_i and C_j : detected corners; black line: detected edges; gray area: matching mask; (a) an example with matching edge pixels $FR_{ij} = 0.96$; (b) an example without enough matching edge pixels $FR_{ij} = 0.54$.

The “fill-ratio” of two corners can also be understood as a measurement of the deviation between the straight line connecting the two corners and the reference edge. We first calculate four “fill-ratio” values: FR_{12} , FR_{23} , FR_{34} , and FR_{41} of each door-corner candidate. If the four “fill-ratio” values are larger than a threshold $FRThresL$ and the average value of the four “fill-ratio” values is larger than $FRThresH$, then the door-corner candidate is determined to correspond to a real door in the image. To handle occlusions, the “fill-ratio” values of two corners that are both formed by two endpoints of open contours (see Figure 2(b, c)) are set to $FRThresH$. Sometimes, more than one door-corner candidate surrounding with a door frame matches with the edge map. In fact, they all represent the same door frame. So, if the overlapped area of two detected doors is large enough, the two detected doors would be merged as one door.

3.4 Distinguishing Doors from Large Convex Objects

Depth information is a significant reference for indoor object recognition. However, a single camera cannot provide depth information. Existing monocular camera-based door detection methods cannot differentiate doors from large object objects with

door-like shape and size, such as bookshelves and cabinets. Although the use of a laser range finder, sonar, or binocular/trinocular camera can solve this problem, it suffers high-cost and high-complexity of the detecting system, which is unsuitable for the wearable indoor navigation aid we envision. Here we propose a novel and simple algorithm by combining lateral information of the detected door to obtain the relative depth information, which would be further used to distinguish doors from other door-like objects. As shown in Figure 4, the concave and convex objects demonstrate the laterals in different positions with respect to the frame. In Figure 4(a), the elevator door, a concave object relative to the wall, demonstrates its lateral ($C_1-C_2-C_5-C_6$) located *inside* of the frame's (C_1-C_2). In Figure 4(b), the bookshelf, a convex object relative to the wall, demonstrates its lateral ($C_4-C_3-C_5-C_6$) located *outside and adjacent* to the frame (C_3-C_4). The observation is true under different perspective scenarios. Such differences are used to infer the relative depth information of an indoor object to distinguish doors from other door-like objects (e.g. bookshelves, cabinets, etc.)

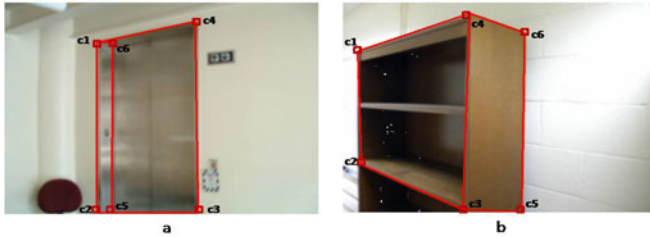


Fig. 4. Laterals of concave and convex objects. (a) A concave object demonstrate its lateral ($C_1-C_2-C_5-C_6$) located inside of the frame. (b) A convex object demonstrates its lateral ($C_4-C_3-C_5-C_6$) located outside the frame.

4 Experimental Results

To validate the robustness and generalizability of our method, we collected a database which contains 203 images of 210 doors from a wide variety of environments. The database includes doors with different colors and texture, elevators, open doors, glass doors, bookshelves, and doors captured with different viewpoints, light changes, and occlusions. Furthermore, based on the complexity of backgrounds, intensity of deformation and occlusion, as well as the changes of illuminations and scale, we categorize the doors into three groups: *Simple*, *Medium*, and *Complex*. The resolution of images used in our experiments is 320×240 . The proposed algorithm achieves accuracy at detection rate of 91.9% with a false positive detection rate of 2.9%. The detail true positive detection results for each category are presented in Table 1. Figure 5 shows some examples of the detected doors from different environments. The first row shows the “*Simple*” examples with simple background. However, it includes a wide variety of conditions of different illuminations and occlusions. The second row shows the “*Medium*” examples with somewhat more complicated background. In some cases, there are multiple doors in a single image. The third row illustrates the “*Complex*” examples with very complex background. This category also includes glass doors and open doors.

Table 1. Door detection results for groups of “Simple”, “Medium”, and “Complex”

Data Category	Number of Images	Number of Doors	True Positive Rate	False Positive Rate
<i>Simple</i>	58	58	98.3%	1.7%
<i>Medium</i>	91	94	91.5%	1.1%
<i>Complex</i>	50	58	86.2%	6.9%
Total	203	210	91.9%	2.9%

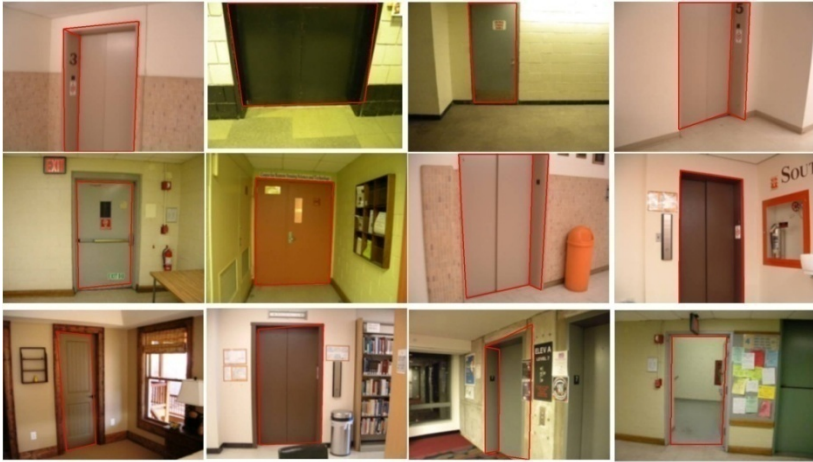


Fig. 5. Examples of successfully detected doors in different environments. The first row shows the “Simple” examples with simple background but include illumination variations and occlusions. The second row shows the “Medium” examples with more complicated background. The third row illustrates the “Complex” examples with very complex background include glass doors and open doors.

5 Conclusion and Future Work

Our goal is to develop a computer vision-based door detection algorithm for navigation and wayfinding to help blind persons independently access unfamiliar indoor environments. In this paper we have proposed a geometric door model that contains only lines and corners of a doorframe. Unlike existing algorithms using the Hough Transform, we develop a new matching method which combines corners and edges. Since our method does not depend on color and other appearance features, it is very robust and can detect doors in unfamiliar environments with variations of color, texture, occlusions, illumination, scales, and viewpoints. This research has the following significant impact: (1) It significantly enriches the study of indoor object detection and leads to significant improvements over existing methods; (2) It provides new strategies and technologies for blind and visually impaired persons to access unfamiliar indoor environments; and (3) It can potentially benefit many other important areas including scene understanding, robot navigation, autonomous systems, etc.

Our future work will focus on detecting and recognizing more types of indoor objects and incorporating context information to improve indoor wayfinding for blind people. We will also address the significant human interface issues including auditory displays and spatial updating of object location, orientation, and distance. With real-time updates, blind users will be able to better use spatial memory to understand the surrounding environment.

Acknowledgments. This work was supported by NSF grant IIS-0957016 and NIH grant EY017583.

References

1. Anguelov, D., Koller, D., Parker, E., Thrun, S.: Detecting and modeling doors with mobile robots. In: Proceedings of the IEEE International Conference on Robotics and Automation (2004)
2. Canny, J.: A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence PAMI-8*, 679–698 (1986)
3. Chen, Z., Birchfield, S.: Visual Detection of Lintel-Occluded Doors from a Single Image. In: IEEE Computer Society Workshop on Visual Localization for Mobile Platforms (2008)
4. Cicirelli, G., D’orazio, T., Distante, A.: Target recognition by components for mobile robot navigation. *Journal of Experimental & Theoretical Artificial Intelligence* 15(3) (July-September 2003)
5. Giudice, N., Legge, G.E.: Blind navigation and the role of technology. In: Helal, A.A., Mokhtari, M., Abdulrazak, B. (eds.) *The engineering handbook of smart technology for aging, disability, and independence*. Wiley, Hoboken (2008)
6. He, X., Yung, N.: Corner detector based on global and local curvature properties. *Optical Engineering* 47(5) (2008)
7. Hensler, J., Blaich, M., Bittel, O.: Real-time Door Detection Based on AdaBoost Learning Algorithm. In: International Conference on Research and Education in Robotics, Eurobot (2009)
8. Hough, P.V.C.: Method and means for recognizing complex patterns. U. S. Patent 3, 069 654 (1962)
9. Kim, D., Nevatia, R.: A method for recognition and localization of generic objects for indoor navigation. In: ARPA Image Understanding Workshop (1994)
10. Munoz-Salinas, R., Aguirre, E., Garcia-Silvente, M., Gonzalez, A.: Door-detection using computer vision and fuzzy logic. In: Proceedings of the 6th WSEAS International Conference on Mathematical Methods & Computational Techniques in Electrical Engineering (2004)
11. Murillo, A.C., Kosecka, J., Guerrero, J.J., Sagues, C.: Visual door detection integrating appearance and shape cues. *Robotics and Autonomous Systems* (2008)
12. National Research Council, Electronic travel aids: New directions for research. Working Group on Mobility Aids for the Visually Impaired and Blind, ed. C.o. Vision. National Academy Press, Washington (1986)
13. Stoeter, S.A., Mauff, F.L., Papanikolopoulos, N.P.: Realtime door detection in cluttered environments. In: Proceedings of the 15th IEEE International Symposium on Intelligent Control (2000)

Development and Installation of Programmable Light-Emitting Braille Blocks

Makoto Kobayashi and Hiroshi Katoh

Tsukuba University of Technology, Japan

koba@cs.k.tsukuba-tech.ac.jp, katoh@k.tsukuba-tech.ac.jp

Abstract. A set of programmable light-emitting Braille blocks which comprise of LEDs were developed. They paved on the approach in front of the main building in the campus of Tsukuba University of Technology. Originally the Braille blocks are useful for the blind, but it also is available for the low-vision. With light-emitting functions, Braille blocks can be good walking guide for them even in the night. These blocks have each ID number and its luminance can be controlled in eight levels. Most remarkable feature is that the program which controls illuminating timing and the brightness is rewritable. By that feature, these blocks became an emergency guide system by changing the animation patterns.

Keywords: Braille blocks, LED, Low-vision, emergency guide system.

1 Introduction

For the persons who have visual impairment, walking guide is general requirement. To fulfill the requirement, many high-technology solutions including wearable devices[1] were proposed. On the other hand, facilitating “Braille blocks” is still good solutions to it. These “blocks” are textured paving blocks especially for the blind persons. The only thing they have to do is following the blocks, without carrying a special hardware, preparing downloaded information, catching GPS signals. In this paper, we start with introducing an environment of the Braille Blocks in Japan.

It is unknown fact that Japan has so many “Braille blocks” paved in the city. When you stay few days in Japan, you can easily find these yellow paving blocks everywhere people walk, not only in a station square, but also on the side walk of the street, even on corridors in some public buildings. The Braille blocks had invented originally by Mr. Miyake who lived in Okayama Prefecture in 1965. Two years later, 230 blocks were paved around a school for the blind in Okayama. After that, the blocks were spreading everywhere. Though there are several arguments that these blocks are not useful for wheelchair users, they became already necessary infrastructure for the blind persons living in Japan.

The blocks have two types of forms to distinguish its meanings. One of them is named “directional block” which contains four straight bulge lines. It indicates a direction people should walk to. The other type is named “warning block” which contains 25 dots arranged in matrix of five-by-five. It indicates that the position is an end

point or an intersection. The warning block also used for the start and end point of a stair or an edge of a platform of a railway station. The precise regulations of blocks are settled in Japanese Industrial Standard (JIS) in September 2001. The standard instituted the number of dots and lines, the shape at cross section, how to pave them, etc. Moreover, there is also guideline book [2]. An example of Braille blocks paved on a side walk is shown in Figure 1.

These Braille blocks are mainly for the blind person mentioned above though, it is also available for the persons who have low-vision hence its vivid color. However, the color effect is weakened in the night. Therefore, few companies in Japan produce and distribute light-emitting Braille blocks equipped with several LEDs. We thought the light-emitting Braille blocks have possibility to improve a night environment for low-vision students in our university. Then we attempted to develop a new type of light-emitting Braille blocks collaborating with one of such companies. Developed blocks were installed in front of the main building and dormitory of our campus.



Fig. 1. Braille blocks paved on a side walk in Japan. The blocks have two types, one of them is “directional blocks” which indicate its direction and the other one is “warning blocks” which indicate end point or intersection point or a kind of caution.

2 Design of Programmable Light-Emitting Braille Blocks

The collaborated company has already produced and distributed a flashing “warning blocks,” which is comprised of high-intensity LEDs and electrical circuit. The LEDs illuminate whole surface of the warning block. Based on their technology and the product, we developed a new type of “directional blocks”, which were designed such that their flashing patterns and brightness can be controlled by a certain program.

At first, we discussed the positions of LEDs. Then we designed as only two of centered straight lines were illuminated, not a whole surface. Because when the whole surface is illuminated, it will be difficult for partially sighted person to distinguish

between warning and new directional blocks. In addition to it, we decided to pave them so as to have several intervals. After all, the illuminated directional blocks have totally different appearance compare with the warning blocks in the night.

Secondly, we considered about the luminance levels and designed it as divided by eight levels. This function enables the blocks to make fine control luminance. Before developing the programmable directional blocks, we had paved 35 light-emitting warning blocks on the edge of stairs in front of our buildings. They are not programmable but can be changed flashing intervals. However, low-vision university staffs and students commented that they had felt the surrounding area would be dark when the flashing light disappeared. It causes by a latency of adaptation of darkness. Therefore, we designed the new directional blocks as to have function of dimmer control.

Thirdly, we discussed how to realize the programmable functions. To make animated illumination patterns by certain program, these blocks have microprocessor and refreshable memory so as to hold the program, and they are connected by data bus cable buried in the ground. The blocks also need an interface to implement a program. We prepare an infrared port for it. And the blocks are designed as to have group number and identical number which are rewritable. Using these numbers, a program determines which blocks should be illuminated. If same group number assigned several blocks in advance, program can switch on multiple blocks simultaneously.

3 How to Program the Blocks

The program of the animated pattern is written in Comma Separated Value (CSV) file format on a computer in the beginning. The CSV file has several records which are composed of several fields. These fields are; a control code, a group number or an



Fig. 2. Left photograph shows a remote control box. A computer writes a certain program to the box through a serial cable. The box stores the program and writes it to a Braille block via infrared. Right photograph shows an overview of light-emitting Braille blocks in the night. Levels of illumination is controlled in eight levels.

identical number of the target Braille block, a parameter like illuminate levels or flashing time, and a waiting time. After making the program, it is transmitted to a remote control box through serial cable. The left photograph of Figure 2 shows the remote control box. After input program to the box, it is carried to the place where the blocks paved on and put the box on a certain block. The block is called “parent block.” Then, the program is written to all blocks using infrared data communication. Writing process is needed for only the “parent block” and program data go through the bus cable to other “child blocks.” After finish writing the program, the animation automatically starts.

4 Implementation as an Emergency Guide

The set of these blocks clearly improved nighttime visibility and it helps to guide low-vision walkers. In the Figure 3, right photograph shows an overview of light-emitting Braille blocks paved in front of the main building. The way to the entrance is clearly indicated even in the night. After constructed them, we paved another set of blocks in front of a dormitory for our students to implement an emergency system. These blocks are slowly animated in usual, however, when an emergency situation happens, prepared program will be written to the blocks through special cable laid from the security guards room. After changing the emergency mode, flashing pattern indicates a right direction and rushes to the safety zone.

The emergency guide system is one example of the blocks we think. Using these Braille blocks, personal guide system which indicates to somewhere user wants to go is able to realize in the future.

Acknowledgement

This works is mainly funded by a “Good Practice Program of Student Support Responding to New Social Needs” and funded by the Ministry of Education, Culture, Sports, Science and Technology, Japan.

References

1. Kim, C., Song, B.: Design of a wearable walking-guide system for the blind. In: Proceedings of the 1st international convention on Rehabilitation engineering & assistive technology (2007)
2. Textured paving block installation guide book, International Association of Traffic and Safety Sciences (2008), <http://www.iatss.or.jp/pdf/other/tenjie.pdf>

FLIPPS for the Blind - An Object Detector for Medium Distances Using Concerted Fingertip Stimulation

Hans-Heinrich Bothe^{1,2} and Sermed Al-Hamdani¹

¹Technical University of Denmark, Oersteds Plads b.352, DK-2800 Lyngby

²Technical University of Berlin, Einsteinufer 17/EN3, D-10587 Berlin

hhb@ieee.org,

s042322@student.dtu.dk

Abstract. The idea of FLIPPS is to design, implement, and evaluate a vibro-tactile device that is composed of a photo diode, vibro-tactile stimulator, micro-controller and power supply. The stimulator is attached to the finger tip and activated when light illumination exceeds a defined threshold. The subject (blind person) receives the reflected light from the objects and, based on brain plasticity principles, can interactively learn to construct a mental image of the objects and of the scenery. The FLIPPS idea is based on sensory substitution theory; here, substituting visual inputs by haptic vibrations.

Keywords: Haptic stimulation, vibro-tactile stimulation, sound source localization, sensory substitution, scene analysis, visually impaired.

1 Introduction

Our idea is to design and build a simple technical system that can be used by visually impaired persons to detect objects in medium distances of a few meters and develop a mental pattern of the actual environment, including static fixed or dynamically moving objects or holes in the ground etc. FLIPPS is a vibro-tactile fingertip stimulator involving several fingers of one or two hands with attached directional sensors. By moving the fingers in a situational but concerted way, the impaired person virtually ‘fingers’ the distant objects interactively and can build his environmental pattern. In order to do so, a learning process is necessary that enables the person to map the vibro-tactile stimulation patterns at his fingers to the environment, and he needs to develop a short-term memory for this mapping.

Theoretically, experiments by M. A. Meredith and B. E. Stein [1] compared neuronal responses in the *superior colliculus* evoked by input from different modalities (multimodal stimulation) against input from each of the single modalities (unimodal stimulation). The results showed a strong response enhancement for multimodal stimulation, the magnitude of which varied with respect to the stimulation level. The closer the input was to the just detectible level, the higher was the multimodal gain. The term brain plasticity was firstly used in neuroscience by W. James in 1896 [2], describing changes in synaptic plasticity due to training. A first mathematical model for such changes was described by D. Hebb in 1949 and described in [3]. M. Fortin,

P. Voss, M. Lasseonde, and F. Lepore [4] then discussed the plasticity of the visual cortex in congenitally blind and showed that the visual cortex is often conducted to non-visual tasks (as, for instance, hearing).

Considering and somewhat generalizing these ideas, we compared the effects and precision of unimodal (vibro-tactile) suboptimal stimulation with multimodal suboptimal stimulation (vibro-tactile plus acoustic) for the tasks i) visual shape recognition and ii) supported auditory sound source detection. In the latter case, the sound was masked with different levels of noise.

2 Technology and Product Description

FLIPPS is an extension to the traditional blind cane that is used by visual impaired people. It intends to help distinguishing light and dark objects or part of objects from each other, independent on distance or illumination. The user then has to learn the principle(s) of how to bind the recognized contrast changes or partial objects and build a realistic scenery by segregating important from unimportant parts. So far, there is a prototype hardware available that is based on sensitive and adaptive directional light sensors, vibro-tactile stimulators, and a processing unit based on a small commercial microcomputer board.

2.1 The Prototype Hardware

The hardware of the prototype is shown in Fig 1; it consists of:

- A glove equipped with four phototransistor sensors; three on the top of the fingers and the fourth on the middle of the hand (Fig 1a), and a vibrator beneath each of the three fingers; a black tube is mounted in front of each of the sensors in order to create highly directional sensitivity.
- A separately worn electronics box (Fig. 1b) includes the electronics, i.e. microcontroller, controllers for the vibrators, a set of programmable gain amplifiers, and the power supply with a 9V battery.
- Furthermore, the electronics box is equipped with a pair of belt hooks and a bootstrap turn-on button (Fig 1b).

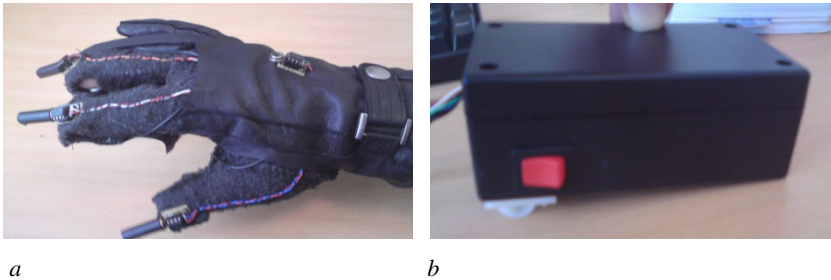


Fig. 1. *a* The glove equipped with the four photodiode sensors and the vibrators; *b* The electronics box equipped with the red on/off button and the belt hooks



Fig. 2. Experimental version of the electronics box: microcontroller board, programmable gain amplifiers, and power supply with a 9V battery

2.2 The Software

One of the fundamental ideas of FLIPPS is that the vibrator activation on each finger is dependent on the light intensity sent by phototransistor sensor on the top of the same finger. The intensity from each sensor is adaptively compared to a sliding threshold value; if it exceeds the current threshold, the corresponding vibrator will be activated. The sensor-activator pairs of all three fingers work independently, so that we can easily vary the number of fingers to use. The threshold is given by an short-time integrated value from the sensor on the back of the hand, which is meant to represent the general lighting situation. Thus, the microcomputer needs to compare the measured values from the first three sensors with those of the fourth sensor. Using a sliding threshold allows the user to quickly adapt to different general lighting situations, for example, to enter a building from outside.

3 Test Protocols

Three different test protocols were ran on artificially blind-folded test subjects in order to show any training progress of the system for purely visual object detection or for visual support of the persons natural abilities of auditory sound source detection; the following is a brief description of these protocols.

3.1 Test Protocol 1

The first goal of this test protocol was to introduce the device to the test subjects and to train them to use the device. The second goal was to test whether the device can be used to recognize simple visible objects by vibro-tactile stimulation in an interactive feedback situation. Simple white objects were presented on a black display screen and the test subject was asked to use the device to recognize the displayed object (see Fig 3a). Based on sensory substitution theories and personal studies, our hypothesis was that by substituting the visual input by haptic input (vibrations), mental images of simple shapes can be constructed and recognized by the blind-folded test subjects.

Before the test started, the objects were shown to the test subjects. We used one line objects as shown in Fig. 3b. In the beginning of the training session, the test subject was then guided and instructed how to use the device. The next phase of the test protocol involves two line objects (see Fig. 3b); the subject was not guided by the trainer. The subjects were asked to name the objects that they recognized.

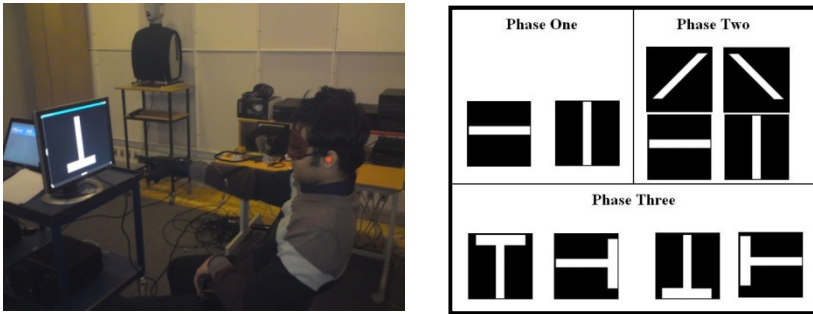


Fig. 3. *a* Test setup: the blindfolded test subject is positioned in front of a black screen equipped with the FLIPPS. *b* Visual objects used in the three phases of the experiment.

3.2 Test Protocol 2

The goal of the second test protocol was to determine quantitatively to what extent the FLIPPS system is able to support intact auditory abilities of the test subjects in noisy environment, i.e. compare the hit rates of suboptimal unimodal- and multimodal stimulation. Unimodal stimulation refers here to either acoustic or haptic stimulation, whereas dual stimulation involves both, acoustic and haptic stimulation.

For this purpose, a white target object in front of a complex background image was projected onto a wide screen ahead of the test subjects. The target object was embedded in a medium grey surrounding in order to compensate for too drastic changes of the lighting and therefore, too easily detectable virtual edges to either the white object and the complex background. The background image comes along with a non-directional speech-shaped noise signal, which associates strong but unintelligible bubbling. The object was either placed at the right or left side or in center of the screen. Together with the presentation of the visual object, a loudspeaker at the same location but behind the screen produced a characteristic speech signal. The test subjects were asked to determine the position of the target object by using all available senses stimulated by either purely binaural auditory, interactive haptic, or by auditory and haptic input. The acoustic signal was masked by noise of a level between 2dB from the just detectable auditory threshold. Our hypothesis was that the dual suboptimal detecting/hit rate was significantly better than corresponding hit rates for unimodal audio or haptic suboptimal stimulation.

3.3 Test Protocol 3

The goal of this test is to study brain plasticity; e.g. to quantitatively determine to what extent the test subjects are able to adapt to stimulation by the FLIPPS system and map complex sceneries to recognizable internal patterns. The three test subjects with the highest hit rates in test protocol 1 were tested (according to the guidelines of test protocol 1) for five subsequent days with a total duration of the test of one hour per day. The daily average hit rates across test subjects were calculated and compared.

A short-term plasticity effect was expected, originated from some prospective functional modifications of the neuronal connectivity. The presented visual objects were letters with two or three lines, namely taken out of the set {L, T, F, H, N, Z}.

4 Results

4.1 Test Protocol 1

In Fig 4 and 5 show the temporal average across trials before announcing the decision verbally, and the average hit rate across trials of the individual test subjects. Fig. 4 demonstrates that the test subjects used in average between 15 and 25 seconds for checking-out the objects by ‘fingering’ it and coming to a conclusion. With the test proceeding. Fig. 5 associates the final average hit rates across trials to values between 68 and 98%, which is significantly higher than chance level (25%). All test subjects were able and capable to use the FLIPPS system for visual-haptic sensory substitution at the end of test phase 1, i.e. ready for the more complicated but still unimodal test phase 2.

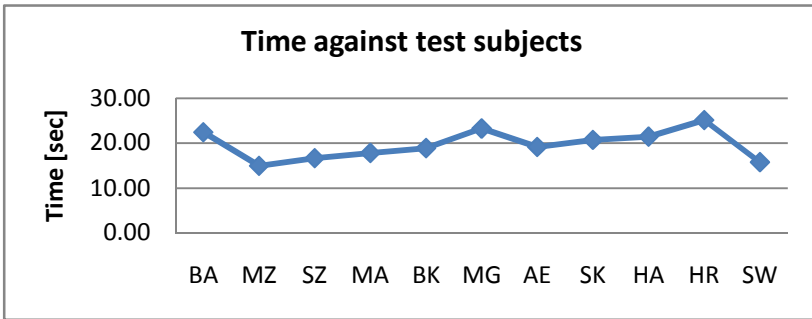


Fig. 4. Average experimenting and detection time across trial with values between 15-25 [sec]

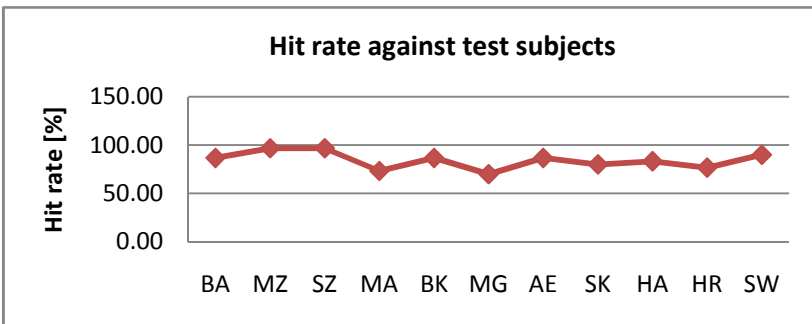


Fig. 5. Hit-rates across trials versus test subjects; shows that the hit rate is significantly larger than the chance rate (25%). This is related to visual-haptic sensory substitution.

4.2 Test Protocol 2

Fig 6. shows the average hit rate across trials on day 1 for unimodal (audio or visual) or multimodal (audio-visual) stimulation. All subjects in all tests (but BA for pure audio stimulation) showed a recognition capacity higher than chance level (33%). For the majority of the test subjects, multimodal stimulation yielded significantly higher average hit rates than unimodal stimulation, which was expected and which is in accordance with [1]. In most cases, multimodal stimulation exceeds the unimodal hit rates significantly. This must be seen as an effort of the FLIPPS system.

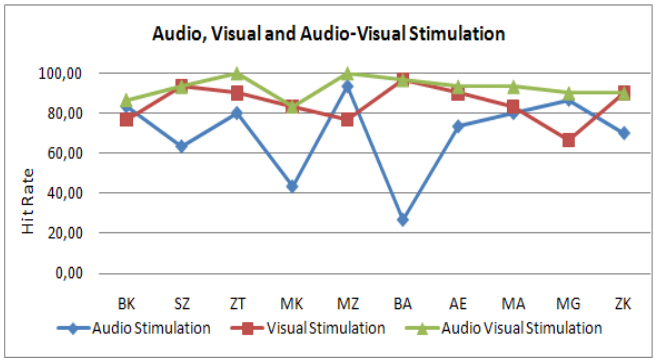


Fig. 6. Improvement of average hit rates across trials

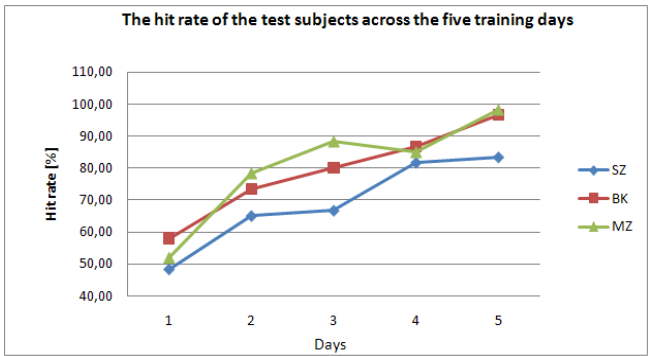


Fig. 7. Average hit rate across trials for audio (diamond), visual (square) and audio-visual (triangle) stimulations

4.3 Test Protocol 3

In order to determine prospective adaptation or learning effects, we asked three persons from test protocol 1 to join an adaptation test for five days, which lasted one hour per day. The average hit rates across trials for each test person and day (see Fig. 7) show improvement between 34 and 46% during the course of the test, which

we relate to the learning effect. The rapid increase during the short time with a limited training duration per day indicates that the FLIPPS prototype system has a high potential for becoming a helpful device for visually impaired persons, even though the present tests were performed in relatively simple lab situations, where the test subjects could calmly sit on a chair and take their time for necessary decisions. It still has to be investigated how visually impaired persons respond to the system and how they react on spontaneous environmental changes detected by the FLIPPS system; this next step is currently in progress.

5 Conclusions

From our results, we conclude that it is possible to construct a mental image of the environment using the FLIPPS system; the related stimulation pattern is yet relatively simple (e.g., objects of a pre-given size with a few lines). However, the rapid improvement of the test results during only a short time also indicates that blind folded subjects may also learn to construct more complex images, which hopefully also holds for visually impaired persons.

Comparing the hit rates from the uni- and dualmodal suboptimal stimulation shows that visual-haptic or auditory-haptic stimulation generally yields a superior recognition against unimodal stimulation. We can hence conclude that the FLIPPS prototype system combined with the natural hearing ability improves object detection efficacy significantly.

Finally, the results of the five days training show a significant improvement of the detection of simple 2-3 lines objects, even though we have only tested three test subjects up to now. We thus hope that our FLIPPS system will become a valuable technical device for visually impaired persons.

Acknowledgement

We like to thank Ron Kupers at Panum Institute and Michael Larsen at Glostrup Hospital, both University of Copenhagen/DK, for valuable cooperation concerning the design of the test protocols and general support, and Hermann Gebert from Technical University of Berlin/D for help with electronic circuitry and mc programming.

References

1. Meredith, M.A., Stein, B.E.: Visual, auditory, and somatosensory convergence on cells in superior colliculus results in multisensory integration. *J. Neurophysiol.* 56, 640–662 (1986)
2. James, W.: *Principles of psychology*, vol. 2. MacMillan, London (1890)
3. Dayan, P., Abbott, L.F.: *Theoretical neuroscience, computational and mathematical modeling of neural systems*, pp. 281–330. The MIT Press, Cambridge (2001)
4. Fortin, M., Voss, P., Lassonde, M., Lepore, F.: Sensory loss and brain reorganization. *Med. Sci. (Paris)* 23(11), 917–922 (2007)

Camera Based Target Acquisition Augmented with Phosphene Sensations

Tatiana G. Evreinova, Grigori Evreinov, and Roope Raisamo

TAUCHI Computer-Human Interaction Unit
Department of Computer Sciences
FIN-33014 University of Tampere, Finland
{e_tg, grse, rr}@cs.uta.fi

Abstract. This paper presents the results of evaluation of the user performance in the target acquisition task using camera-mouse real time face tracking technique augmented with phosphene-based guiding signals. The underlying assumption was that during non-visual inspection of the virtual workspace (screen area), the transcutaneous electrical stimulation of the optic nerve can be considered as alternative feedback when the visual ability is low or absent. The performance of the eight blindfolded subjects was evaluated. The experimental findings show that the camera-based target acquisition augmented with phosphene sensations is an efficient input technique when visual information is not available.

1 Introduction

Human-computer interaction strongly relies on the interface structure and feedback cues that were exclusively designed by sighted people for sighted users based on visual concepts and metaphors. To improve the design of virtual workspace for sensory challenged people, the perceptual, cognitive and motor strategies they employ to coordinate motor activity within the personal haptic-perceptual space have to be considered. The interface should be transparent to the user even in the absence of visual feedback, therefore the spatial locations which cannot be accessed with regular tools and techniques for pointing and selection have to be arranged regarding their dominance, priority content, accessibility and mnemonics [3].

The conventional input devices such as mice or joystick can provide the sensory impaired user with some kinesthetic feedback depending on their design and the interaction style. However, for visually impaired and blind users graphic input should rely on non-visual feedback signals. Such a graphic-like input technique might be acceptable even in the absence of visual feedback, but direct manipulation should be augmented with alternative continuous feedback cues. In addition, the hand input might not be the best choice to facilitate access to information for blind user considering the fact that the user is compelled to produce accurate movements in exact vertical or horizontal directions of the pointer with particular speed and acceleration in the absence of visual feedback [1].

Fortunately, video-based input in a near field for the sensory challenged PC user is powerful and promising technique. In the video-based input [1, 2, 7, 8, 11], the user

gaze direction can be measured and tracked as a displacement of the facial feature regarding the capture window. Thus, head movements might be translated into a position of the virtual pointer. Blindfolded subjects [2] demonstrated twice better performance in the target acquisition task with camera-joystick vs. manual joystick.

Our underlying assumption is that during non-visual inspection of virtual screen (workspace), the transcutaneous electrical stimulation of the human optic nerve can be considered as alternative feedback when the visual ability is low or absent.

It was discovered already by Otfried Foerster in 1929 that applying electrical stimulation to the retinal area of the eye evoked visual sensations of points, spots, clouds, and bars of colorless or colored light (phosphene phenomenon) in humans. Later, several surgical and non-invasive attempts undertaken on electrical stimulation of occipital cortex and optic nerve of visually challenged and blind patients [9] have proven that the visual sensations are not only easily perceivable and distinguishable from each other [10], but can also stimulate partial recovering the visual sensation in blind subjects such as the possibility of light perception, motion detection, and even discrimination of simple shapes. At that, the induced phosphene sensations structured within accepted safety limits discussed in Shannon [12] do not cause any clinically visible cellular damage or vision loss.

In the previous studies [4, 5, 6], we have demonstrated that properly structured phosphene-based cues, delivered through the electrodes mounted into frame of spectacles and having a good contact with temples can give a rise to a clear, perceptually controlled phosphene sensations well distinguished visually and associated with different meaning. Phosphene-based cues could be attractive candidates for providing supplementary information or feedback regarding the features of different objects (distance and relative location, elevation etc.) and regions of the visual scene, while the resources of visual cortex would process the induced modulation of neural activity of the retina or/and optic nerve.

The goal of this study was to evaluate the blindfolded user performance in the target acquisition task in the presence of phosphene sensations augmenting video-based head input in the absence of visual feedback. Navigation within screen space was augmented with phosphenes. The intensity of phosphenes was gradually changing depending on a distance to targets. The phosphene sensations were provided through electrodes mounted on a frame of spectacles by stimulating transcutaneously optic nerve.

2 Method

2.1 Participants

Eight right-handed volunteers (four females and six males) from the staff and students at the University of Tampere were involved in the study. Four of the subjects were males and four were females. Eight of the participants had a low vision corrected with glasses below 20/100. The age of the subjects ranged from 28 to 54 years with a mean age of 36. The subjects used a computer on a daily basis, reporting 7 to 10 hours of usage per day. Four of the subjects had limited experience with the camera-mouse pointing technique used in the presence of visual feedback. Two of participants have had prior experience with use of electro-tactile display techniques transcutaneously

stimulating human optic nerve. None of participants had health or skin sensitivity problems that could hamper their participation in the experiment.

2.2 Experimental Setup

Asus W7F laptop (Intel Core Duo T2400, 1.83 GHz) was used in the experiment. Integrated web-camera ASUS WebCam with resolution of 1.3 MP was set to the black and white image mode with horizontal mirror and a frame rate of 30 fps. Mouse movements were translated into a screen cursor location in a conventional way with matching directions (up, down, left, right), and augmented with phosphene-based guiding signal of three different intensities (for more description see section 2.3) by means of the eGlasses implemented in our earlier studies. The eGlasses had non-transparent glasses and the electrodes mounted into a frame of spectacles so that they could have good contact with temples (Figure 1).

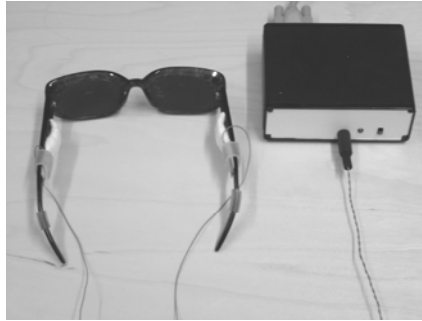


Fig. 1. The eGlasses and electro-tactile unit used in the experiment

The eGlasses were connected with the electro-tactile unit and manipulated by means of the software implemented in our previous studies [4-6] under Microsoft Visual Basic 6.0 which was used to produce predefined parameters of the phosphene such as number, intensity, polarity, and delay between the electrical pulses.

The camera-mouse module implemented in previous research [5] was making use of Microsoft Video for Windows (VfW) and API functions. The test program was written in Microsoft Visual Basic 6.0 under Windows XP. The camera-mouse module was used for relative pointing and has an acceleration factor of 2.4 and velocity factor of 1.7 employing a sigmoid-based transfer function for better controlling the cursor in a head gestures-related selection task.

2.3 Phosphenization

The phosphene-based guiding signal consisted of one burst with duration of positive and negative pulses of 80 ms each and with a delay of 5 ms between pulses. The resulting duration of signal was about 170 ms. The signal elicited the intensive sensation of the bright white light spot.

To guide the target acquisition process, the intensity of phosphene-based guiding signal was modulated by changing their parameters depending on the distance to the target. The stimulation parameters of guiding signal were modulated in three ways:

1. When the distance to target location (D_{trg}) was greater than 200 pixels, the intensity I_1 of the guiding signal was modulated by the delay between electrical pulses of positive and negative polarity within the burst and was calculated as follows:

$$I_1 = |(((D_{trg} - 200)/120) + 1) \times 30| \quad (1)$$

The guiding signal of intensity I_1 evoked in the retinal area of visual analyzer the sensation of *slowly flickering lights of white color*.

2. To fulfill the condition $100 \text{ pixels} < D_{trg} < 200 \text{ pixels}$, the intensity I_2 of the guiding signal was modulated by the duration of the electrical pulses of positive polarity within the burst and was calculated as follows:

$$I_2 = |(((D_{trg} - 100)/20) + 1) \times 5| \quad (2)$$

The guiding signal aroused the sensation of *mildly flickering lights*.

3. If D_{trg} was less than 100 pixels, the intensity I_3 of the phosphene was modulated by the duration of the electrical pulses of negative polarity within the burst. The guiding signal elicited the sensation of *rapidly flickering lights*.

The number of guiding signals of different intensities (N_1, N_2, N_3) spent by the subjects during experiment indicated how the subjects navigated regarding the target to be captured, and allowed to estimate their performance.

2.4 Procedure

The subjects were tested individually in the usability laboratory at the local university. Prior to data collection, the experimenter briefly introduced the software setup, camera-mouse functionality and the instructions concerning the overall testing procedure. The subjects, who wore glasses, were asked to take them away. Before using the eGlasses, the electrolyte gel was applied to the temples of the participants to reduce the impedance of the contact between electrodes and the skin.

The task of the subjects consisted of acquiring targets augmented with phosphene-based guiding cues. Detecting the target was accomplished by means of the virtual pointer being manipulated by directional head movements (blindfolded “gazing” through eGlasses) within the test area of the software application ($285 \times 180 \text{ mm}$). The targets (spots) were arranged in a rectangular grid of 3 rows by 3 columns. The size of rectangular targets was set to $15 \times 15 \text{ mm}$ and the distance between the centers of adjacent spots was set to 70 and 42 mm along vertical direction. Such a layout was being inspected from the distance of 400 mm with the camera-mouse.

Every time the virtual cursor entered the target zone, a short sound was played to notify the subject that the target was successfully detected. To avoid accidental crossing the target, the pointer should stay within the target not less than 50 ms, and the next target to be located was consequently exposed to the participant upon the confirmation of target selection. The targets were selected randomly and presented all

nine of them at once per test trial. The Randomize function was used where appropriate to avoid similar random sequences. As the starting position of pointer and the position of the next target to be located were varying randomly, the routes to navigate were different for each subject and each trial. Upon detection of all the targets, sound signal indicating the completion of the trial was heard, and the subject could start new test trial by pressing the space bar on the laptop keyboard when s/he felt ready.

The participants were instructed on how to capture the targets accompanied by phosphenes relying on their perception of the flickering lights intensity. After that, they were given one warm-up trial to familiarize themselves with the features of the blind target acquisition and principles of using the navigational phosphene-based cues.

Each subject completed five test sessions with no more than one session per day. Each session included fifteen trials, but the data of the first trial were excluded from analysis. Each trial consisted of nine target acquisition tasks. The subjects could rest as desired between the trials. One session took approximately two hours. After training, all the data such as the number of phosphenes activated during movement along trajectory to the target position, the distance to the target being acquired, and the time spent by the subjects to accomplish the target acquisition task, were automatically collected and stored in a log file for further analysis.

3 Results

Figure 2 illustrates how the number of guiding cues N_1 , N_2 , and N_3 varied throughout the five test sessions being averaged over nine targets, that is, independently on the target distance. To ease the target detection process, the participants were very careful about the choice of the navigation route. Once the subjects heard the short sound indicating that the target was exposed, the participants had to discriminate and follow changes in the intensity and frequency of the sensation of flickering lights until they would feel that the guiding signal starts to flash rapidly and then they promptly detected the target.

In the first test session, the subjects were still novices in estimating variations in the intensity and frequency of the guiding signals. They made promiscuous forward or backward moves toward the target position, and therefore the total number of flickering lights was great: $N_1 = 114$, $N_2 = 93$, and $N_3 = 21$ with intensities proportional to the distance (see eq. 1, eq. 2).

As the participants began to strongly rely on changes in intensity and frequency of the guiding signals, the total number of flickering lights significantly decreased: $N_1 = 53$, $N_2 = 44$, and $N_3 = 8$ (see Fig. 2).

The paired-samples t-test result for the total data averaged over all target positions detected by all participants revealed that there was a statistically significant difference between the average number of the guiding signals accompanying the navigation to target in the 1st and 5th test sessions: $t(8) = -11.75$, $corr. = 0.989$, $p < 0.0001$.

Figures 3 and 4 show that gradual acquiring the experience in estimating phosphenes (the changes in intensity and frequency) resulted in better performance in the end of the experiment. In the 1st test session, the players were required of about 5250 ms to detect nine targets while in the 5th test session they have spent about only 458 ms to complete the targets acquisition task. The targets 1, 3, 7 and 9 remotely

located from the center of navigation field in the 1st test session required more efforts and consequently greater number of guiding cues than the center target 5 and neighborly located targets 2, 4, 6 and 8. The same tendency was observed in the 5th test session with the targets 2, 3, and 4 having the greater distance from the center of navigation field among the other targets.

The ANOVA analysis of variance revealed a significant difference between the grand mean target acquisition times spent by the participants in the 1st and in the 5th test sessions ($F = 45.82, p < 0.001$).

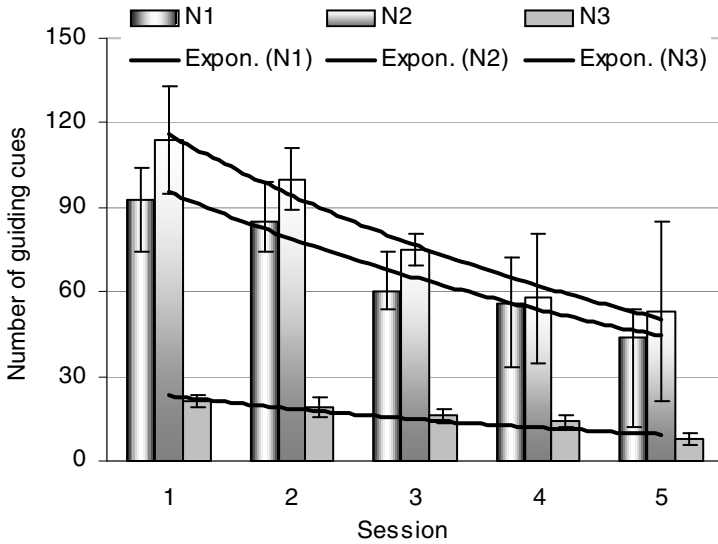


Fig. 2. The number of the phosphene-based guiding cues N_1 , N_2 , and N_3 used to detect the target location and averaged over nine targets throughout the five test sessions

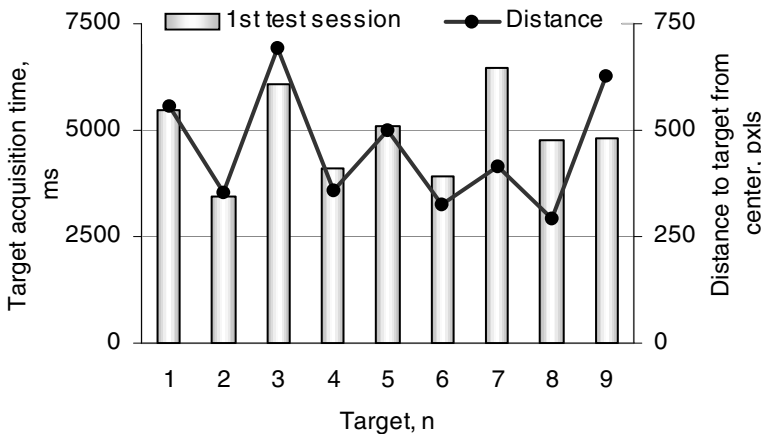


Fig. 3. The grand mean target acquisition times spent by the participants to detect nine targets in the 1st test session

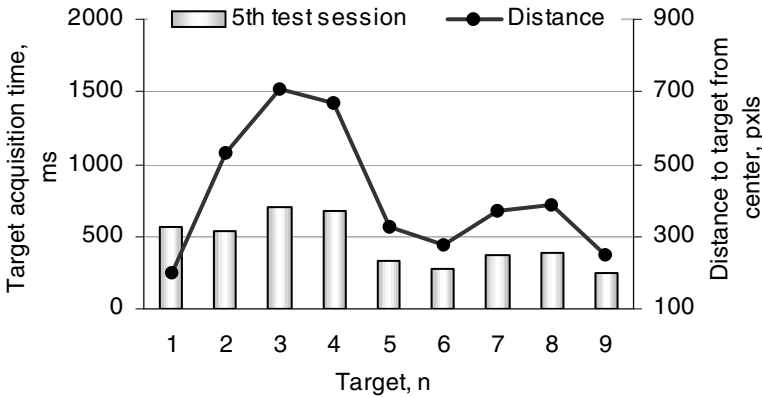


Fig. 4. The grand mean target acquisition times spent by the participants to detect nine targets in the 5th test session

4 Conclusion

We evaluated the user performance in the target acquisition task using camera-mouse input technique augmented with phosphene-based signal of different intensities provided through electrodes mounted on a frame of the spectacles and stimulating transcutaneously the optic nerve. The performance of the participants was evaluated in terms of the number of phosphenes of different groups of intensity values accompanying head movements to navigate to the target, and the time spent by the participants to accomplish the target acquisition task. The experimental results revealed that gradual acquiring the experience in estimating the changes in intensity and frequency of the phosphene-based guiding cues by the subjects resulted in more than twice better performance to the end of the experiment. The target acquisition times spent by the participants have also significantly decreased by about tenfold to the end of the experiment. The encouraging experience acquired within simulated conditions can be applied for designing camera-based input techniques with alternative continuous feedback to facilitate the information access for blind and visually impaired users. The future work will be focused on thorough assessing the personal adaptation features to transcutaneous electrical stimulation and developing new interaction scenarios incorporating the phosphene-based feedback cues in assisted techniques.

Acknowledgments. This work was financially supported by Academy of Finland (grants 115596 and 127774).

References

1. Betke, M., Gips, J., Fleming, P.: The Camera Mouse: Visual Tracking of Body Features to Provide Computer Access For People with Severe Disabilities. *IEEE Trans. on Neural Systems and Rehabilitation Engineering* 10(1), 1–10 (2002)

2. Evreinova, T.V., Evreinov, G., Raisamo, R.: A Camera-joystick for Sound-Augmented Non-Visual Navigation and Target Acquisition – a Case Study. In: Universal Access in the Information Society, vol. (7), pp. 129–144. Springer, Heidelberg (2008)
3. Evreinova, T.V., Evreinov, G., Raisamo, R.: Camera based Head-mouse: Optimization of Template-Based Cross-Correlation Matching. In: Proc. of Int. Conference on Computer Vision Theory, VISAPP 2007, Barcelona, Spain, March 2007, pp. 507–514 (2007)
4. Evreinova, T.G., Evreinov, G., Raisamo, R.: A Comparative Evaluation of Color Blinking and Electro-Tactile Imaging of Hidden Graphs. In: Proc. of the IADIS Int. Conf. on Interfaces and HCI. IADIS 2008, The Netherlands, July 25-27, 2008, pp. 109–117 (2008) ISBN: 978-972-8924-59-1
5. Evreinova, T.G.: Visualization of Math Expressions through Modality-Nonspecific Signals. In: Universal Access in Information Society, October, vol. 8(3) (2009)
6. Evreinova, T.G.: Transcutaneous Stimulation of the Human Optic Nerve For Encoding Math Expressions. submitted to IADIS Int. Conf. on Interfaces and HCI (2010)
7. FaceMOUSE. Product information (2005), <http://www.aidalabs.com/>
8. Jilin Tu, T., Huang, T., Tao, H.: Face As Mouse Through Visual Face Tracking. In: Proceedings of the 2nd Canadian Conf. Computer and Robot Vision (CRV 2005), Victoria, BC, Canada, May 2005, pp. 339–346. IEEE Computer Society Press, Los Alamitos (2005)
9. Humayun, M.S., de Juan Jr., E., Weiland, J.D., et al.: Pattern electrical stimulation of the human retina. *Vision Res.* 39, 2569–2576 (1999)
10. Rizzo III, J.F., Wyatt, J., Loewenstein, J., Kelly, S., Shire, D.: Methods and perceptual thresholds for short-term electrical stimulation of human retina with microelectrode arrays. *Investigative ophthalmology & visual science* 2003 44(12), 5355–5361 (2003)
11. Sako, H., Whitehouse, M., Smith, A., Sutherland, A.: Real-time facial-feature tracking based on matching techniques and its applications. In: Proc. of 12th IAPR Int. Conf. Pattern Recognition. Conference B: Computer Vision & Image Processing, vol. 2, pp. 320–324 (1994)
12. Shannon, R.V.: A model of safe levels for electrical stimulation. *IEEE Transactions on Biomedical Engineering* 39(4), 424–426 (1992)

A Mobile Phone Application Enabling Visually Impaired Users to Find and Read Product Barcodes

Ender Tekin and James M. Coughlan

The Smith-Kettlewell Eye Research Institute, San Francisco, CA
{ender, coughlan}@ski.org

Abstract. While there are many barcode readers available for identifying products in a supermarket or at home on mobile phones (e.g., Red Laser iPhone app), such readers are inaccessible to blind or visually impaired persons because of their reliance on visual feedback from the user to center the barcode in the camera's field of view. We describe a mobile phone application that guides a visually impaired user to the barcode on a package in real-time using the phone's built-in video camera. Once the barcode is located by the system, the user is prompted with audio signals to bring the camera closer to the barcode until it can be resolved by the camera, which is then decoded and the corresponding product information read aloud using text-to-speech. Experiments with a blind volunteer demonstrate proof of concept of our system, which allowed the volunteer to locate barcodes which were then translated to product information that was announced to the user. We successfully tested a series of common products, as well as user-generated barcodes labeling household items that may not come with barcodes.

1 Introduction

The 1D barcode is a ubiquitous system for labeling packaged goods with codes that uniquely identify product information. This type of barcode (which includes the UPC and EAN symbologies, widely used in North America and Europe, respectively) was designed to be read by a laser scanner, which is a standard tool in supermarkets and shops, but there is increasing interest in being able to read barcodes using the most ubiquitous portable device: the mobile phone.

Applications such as the iPhone Red Laser, Nokia's Point and Find and Realeyes3D have recently been released to provide this functionality, but all such products force the user to locate and home in on the barcode using visual feedback from the camera viewfinder. Without such feedback, it is nearly impossible to guide the camera close enough to the barcode to view it with sufficient resolution. As a result, these products are only accessible to people with sufficiently good vision.

Dedicated devices such as the i.d. mate OMNI talking barcode reader (from Envision America), which uses a laser to rapidly scan a product and read any barcode within its field of view, have been designed expressly for blind and visually impaired users, but such products are expensive (e.g. over \$1000 US) and require the use of a dedicated piece of hardware. One possible alternative to such a dedicated device is a system such as Trinetra [3], which uses a small barcode reader that plugs into a standard mobile

phone; however, the barcode reader is still an additional device that makes the system less convenient to use.

A computer vision algorithm enabling a blind person to locate and read barcodes can be implemented on a camera phone, perhaps in addition to other functionality providing access to printed information (such as the KNFB reader, which is a commercial OCR system that runs on a Nokia camera phone). Such a system would be much more convenient than conventional laser scanner technology that requires the user to carry a dedicated device, such as the i.d. mate.

The alternative approach that we describe in this paper is to use a portable computer vision-based system to find a barcode, direct the user with audio feedback to this barcode until it is well-centered in the camera frame, and then decode the barcode. This system is implemented as a software application on an off-the-shelf mobile phone, building on a previous desktop computer version of the system implemented by the authors [5,4]. Once the barcode is decoded it is matched to an on-device database to identify the product (no data connection is required), and the corresponding product information is presented using text-to-speech functionality.

2 Finding and Reading Barcodes with Computer Vision

In this work we specialize to UPC-A barcodes, which are widely used in North America to label commercial products, but it will be straightforward to extend our system to other common symbologies such as EAN (common in Europe).

The computer vision algorithm we have developed has two stages, the first for finding barcodes in cluttered images, the second for reading them when they are viewed at close range, as presented in the system overview in Fig. 1. We note that past work on computer vision algorithms for reading barcodes on mobile phones [2,6] has de-emphasized the important issue of finding the barcode, instead assuming that the user can center the barcode in the viewfinder.

Our application starts in a low-resolution mode for the first part of our algorithm, which is the barcode detection part, so as to run at a reasonable speed. It first determines if there is a piece of a barcode in each video frame it receives. This is accomplished by searching lines of pixels along the 0° , 45° , 90° and 135° orientations, and seeing if there are multiple lines that are approximately perpendicular to the swept lines. Areas with such lines are recorded as candidate areas for further investigation.

The algorithm then focuses on the biggest such verified candidate area if there are multiple such candidates. If this area is too small, or too close to the edge of the image (implying that there may be a yet unseen portion of a barcode), then the user is notified by directional feedback (which can be chosen to be a different tone or voice command) until the candidate area is large enough and relatively well positioned in the frame. We then use a higher resolution version of the barcode area to detect and localize the edges more accurately. If enough edges are found, we use a Bayesian statistical approach, [4], to find the barcode digits that are most likely, based on the edge locations. Multiple slices are used if needed, until we have sufficient confidence about the digit estimates.

Then the detected barcode is matched against the product databases, first the customizable user database and then the full database. If the information for this product is

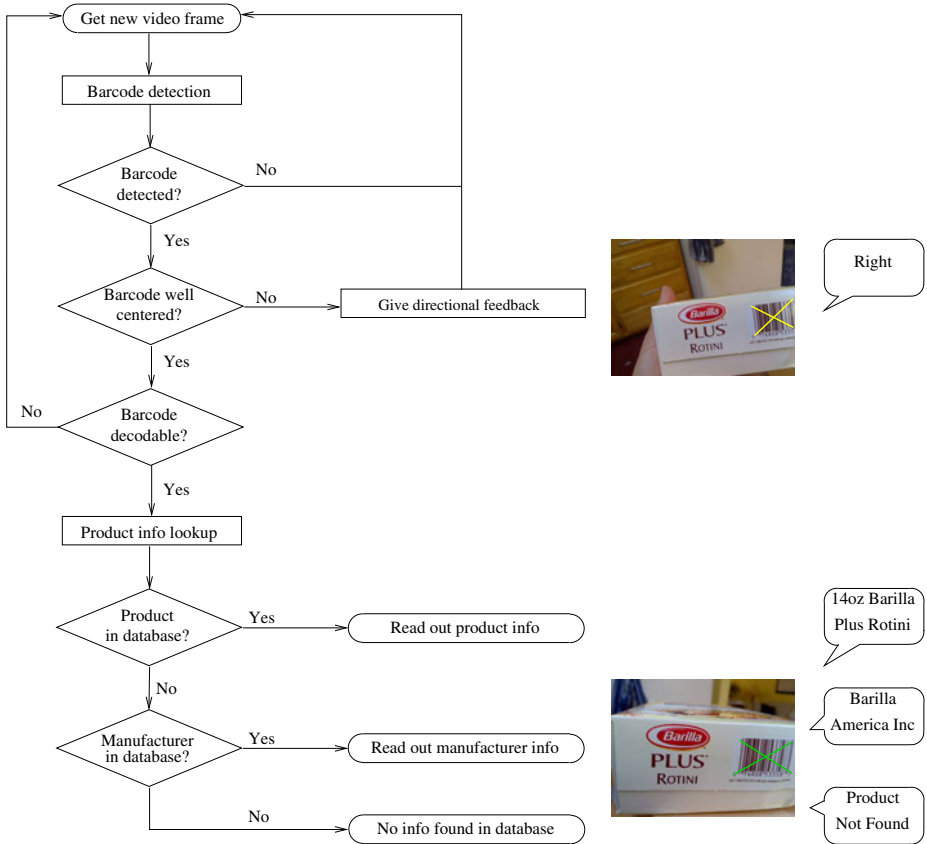


Fig. 1. System overview

found, it is read out using text-to-speech. If we cannot find an entry, then we match the first six digits of the barcode against a manufacturers’ database. If the manufacturer is found, then this information is read out loud; if this information is not available either, the user is notified of the lack of information regarding this product.

3 System Operation and User Interface

The preceding barcode location and decoding algorithm was developed and tested on a desktop computer and then ported to the Nokia N97 camera phone running Symbian C++. The algorithm was implemented to run as quickly as possible on the mobile phone, so that up to a few images could be processed per second. To achieve operation as close to real-time as possible, we had the system run in a video capture mode (in which several images are automatically captured every second).

The system’s user interface guides the user through two stages for each product, first finding and then reading a barcode. The user holds the camera approximately

10-15 cm above the surface of the product, slowly scanning the likely locations of the barcode (see Fig. 2); the system is silent until it finds sufficient evidence for a barcode in its field of view. When a barcode is detected, the system attempts to guide the user to center the barcode in the field of view, using four distinct signals (tones or verbal instructions) to indicate that movement is required to the left, right, up or down. If the barcode is centered but too far to be clearly resolved and read, another tone directs the user to bring the camera closer to the barcode. Once the barcode is sufficiently close, the system continues capturing images of the barcode until a barcode estimate can be determined with likelihood exceeding a threshold, and indicating these attempts with another tone.

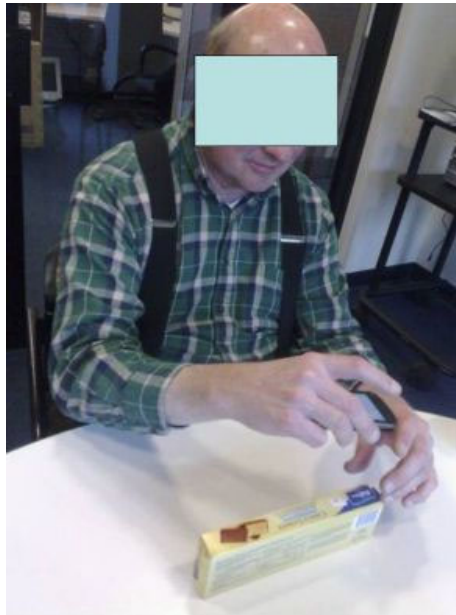


Fig. 2. Blind volunteer subject (with face covered for anonymity) using our system

Finally, when a barcode is read with sufficient confidence, a “successful read” signal is issued, the barcode is looked up in a freely available UPC database [1], and the results of this look-up are read aloud using text-to-speech functionality (built into the mobile phone). We note that the UPC database contains over one million entries (and is augmented with any user-generated barcode entries, if any) but is small enough (under 70 MB total) to be downloaded to the mobile phone’s flash memory. Thus, the system software runs locally on the mobile phone without the need for any data plan to access a database on a remote server, which is an important advantage if the user is in a place without mobile reception.

4 Experimental Results and User Testing

We explained the purpose and operation of the system to a completely blind volunteer subject, and advised him on how best to hold the mobile phone (Fig. 2). Particular emphasis was placed on the importance of moving the camera slowly to avoid motion blur and to ensuring the camera lens was not covered (e.g. by the user's fingers). In addition, we discussed the importance of using prior knowledge of where barcodes are likely to be found, and at what orientation, on different packages to narrow the search. For instance, barcodes tend to be aligned with the edges of a package, and the camera is better able to resolve a barcode when it is aligned to the barcode (or at a 90° angle); on a cylindrical can, the barcode is on the circumference of the cylinder near the flat top or bottom.

We first conducted a brief training session on several barcoded packages, in which we gave advice to the subject as he was using the system; some of the packages in this session that he was able to read included user-generated barcodes, i.e. barcodes that we printed out to label personal items that did not already originally have their own barcodes.

Next, a simple experiment tested the subject's ability to find and read barcodes for ten items using the system without our help. These ten packages were commonplace grocery items in the form of boxes, cans and bottles (separate from the items used in the training session). One item was a bottle of Ibuprofen on which we pasted the correct product barcode, since the bottle lacked its own. The subject was able to locate barcodes on nine of the ten packages but was unable to locate the barcode on a plastic jar, because of false positive detections due to the presence of densely grooved lines on the top of the jar (which resemble barcode lines). Of the nine barcodes that were successfully located, seven were read correctly and one was located but not read; the other was read with two incorrect digits (out of the entire 12-digit barcode digit sequence), but enough digits were correct to determine the product manufacturer (indicated by the first six digits in the sequence).

While our experiments establish proof of concept of our system, there are currently two main problems that need to be solved to make this a truly useful product. The biggest problem is that the Nokia N97 phone we used has a slow and erratic camera autofocus mechanism, which results in some of the images being out of focus, thereby degrading the system's performance. Also, the phone CPU is much less powerful than a standard desktop computer, which makes the system much slower to analyze images than the desktop version [4], and limits the system's ability to quickly rule out false positives. However, both these problems should be ameliorated in newer-generation devices such as the iPhone or the Google Nexus One, which we plan to experiment with in the future.

5 Conclusion

We described a new mobile phone system that enables blind and visually impaired users to find and read barcodes with a mobile phone. The system is implemented as a software application that can be downloaded on an off-the-shelf phone, and it is the first

such system we are aware of that is accessible to blind users that doesn't require the use of any additional hardware. Preliminary experiments with a blind subject demonstrate the system's feasibility.

Future improvements to this system include porting it to more popular and powerful mobile platforms, such as iPhone and Android, improving the speed and accuracy of the computer vision algorithms for finding and detecting barcodes, extensive testing with blind and low vision subjects, and extending the system to handle a greater variety of barcode symbologies (such as EAN).

Acknowledgments

The authors were supported by The National Institutes of Health grant 1 R01 EY018890-01.

References

1. Internet UPC Database, <http://www.upcdatabase.com/>
2. Kongqiao, W.: 1D barcode reading on camera phones. *International Journal of Image and Graphics* 7(3), 529–550 (2007)
3. Narasimhan, P.: Trinetra: Assistive technologies for grocery shopping for the blind. In: *IEEE-BAIS Symposium on Research in Assistive Technologies*, Dayton, OH (April 2007)
4. Tekin, E., Coughlan, J.M.: An algorithm enabling blind users to find and read barcodes. In: *IEEE Workshop on Applications of Computer Vision*, Snowbird, UT (December 2009)
5. Tekin, E., Coughlan, J.M.: A bayesian algorithm for reading 1d barcodes. In: *Sixth Canadian Conference on Computer and Robot Vision*, Kelowna, BC, Canada (May 2009)
6. Wachenfeld, S., Terlunen, S., Jiang, X.: Robust recognition of 1-D barcodes using camera phones. In: *9th International Conference on Pattern Recognition*, Tampa, FL, December 2008, pp. 1–4 (2008)

A Model to Develop Videogames for Orientation and Mobility

Jaime Sánchez^{1,2}, Luis Guerrero^{1,2}, Mauricio Sáenz^{1,2}, and Héctor Flores¹

¹ Department of Computer Science, ² Center for Advanced Research in Education (CARE)
University of Chile
Blanco Encalada 2120, Zip Code 2777, Santiago, Chile
{jsanchez, luguerre, msaenz, hflores}@dcc.uchile.cl

Abstract. There is a real need to have systems for people with visual disabilities to be able to improve their orientation and mobility skills, and especially for children to be able to improve their autonomy into the future. However, these systems must be designed according to available objectives, methodologies and resources, as well as by taking the interests and ways of interacting of the end users into account. This work presents a model for the development of videogame-based applications, which includes differing levels of abstraction and different stages in the design and development of systems that allow for the improvement of orientation and mobility skills for people with visual disability. The feasibility of the model was studied by modeling two videogames for children with visual disabilities.

Keywords: Software engineering model, serious videogames, orientation and mobility, audiogames.

1 Introduction

Spatial relations during the first years of life are established through the ability to situate oneself in space, using different sensory channels and strategies that allow one to assimilate and integrate the information intake from the environment. This information is transformed into meaningful knowledge that we use throughout our lives [2].

In the case of a blind child, in the sensory-motor period of his/her development, aspects such as the permanence of an object is a difficult concept to acquire, given that vision is the main sense that provides information on the presence of objects or people in a certain location [2]. For a child with visual disability to be able to develop this ability, it is necessary to be accompanied verbally during all the movements that he/she makes.

Unlike a sighted child, for whom movement emerges as a result of visual curiosity, a blind child lacks visual experience, such as seeing oneself in the mirror, seeing other people and relating with them, or feeling a visual attraction towards an object [10]. As a child with visual disability does not have the chance to be “attracted” to things, his/her mobility will be diminished, as initially sound is not able to transmit the existence of things that can be touched; although in time and with the positive

development of auditory perception, this can change [10]. As a consequence, the experiences of a child with visual disability as compared to a sighted child regarding the initial stages of life, do not allow the visually disabled child to integrate tactile, auditory and kinesthetic experiences in the same way as a sighted child [7].

As far as orientation, children with visual disability should learn and understand from an early age that to be oriented is to know where one is, and to know the relationship between one's position and other locations in space. They must learn to establish and utilize reference points that allow them to know where they are located at any given time in their navigation [17].

There are several ways to help users with visual disabilities to achieve autonomous navigation with the support of technology. Such technologies can be defined as any item, equipment or system that is commercially acquired, and is modified or adapted in such a way that it can be used to increase, maintain or improve the functional abilities of blind users [13]. Some projects provide different means of interaction for blind users through the use of mobile devices, implementing ways of entry through the use of tactile or voice commands, the output of which is provided through verbal and/or iconic sounds [5][15].

Bradley and Dunlop [3] consider the contextual differences in which the blind user moves about, differentiating between indoor and outdoor systems. The main result includes the fact that with location systems for legally blind people, both technical aspects of the technology and the different ways in which the users codify spatial information according to the context in which they are immersed must be taken into account. Other mobile aid solutions for the orientation and mobility of legally blind users take advantage of the logic of the environments with which these users relate. Sánchez and Maureira [15] present mBN, a system of navigation to be used in a subway station. Without the need of any other device besides a handheld PocketPC, the user can obtain information from the various subway stations and certain aids that allow him/her to enjoy autonomous navigation. Another way to help them to be more autonomous is to train them virtually, to then apply the knowledge attained in the real world [1]. In the studies carried out by Lahav and Mioduser [6] the user's achievement regarding their cognitive representation of the virtually navigated space that they attain, and their ability to apply this representation in carrying out tasks in a real space are examined. The results show the success of the experience, in which the users are able to construct a mental map and then apply it to the real world.

In general, videogames are seen only as tools for entertainment. They can also be used as powerful learning tools [12]. However, although the majority of successful studies and projects in this area refer to sighted children, there is a clear niche for research on the application of videogames in support teaching new skills to blind children.

In this way, for the purpose of this study it is necessary to evolve from a traditional software development process, and adapt it to these new needs (blind children, serious videogames, orientation and mobility learning), thus creating a new development model. Therefore, the objective of our study is to propose a model for the development of videogame-based applications designed to assist the navigation of blind people, based on the adaptation, widening and integration of existing software engineering models and the design and evaluation of mobile applications that support the

orientation and mobility of blind people [16][11] to this new field within software engineering, related to the design and creation of videogames for the education and learning of blind children.

2 Methodology

To design this innovative and specific model for the development of video game-based applications for the navigation of blind people, other models that have already been designed, developed and validated by the authors of this study were used as a reference. These models, however, do not consider the video game factor and the multi-disciplinary nature of the focus for the current study. The first model taken for reuse is related to the design, development and evaluation of mobile applications to support the development of orientation and mobility (O&M) skills in blind people [16]. The second model is related to software engineering, and is used to support in the development of conventional mobile applications for learning. This model allows the developers to consider aspects that are critical to the development of a functional application [11].

3 Model

In order to develop video game-based applications for the improvement of O&M skills in blind users, it is necessary to consider three processes (see Fig.1): (1) Definition of Cognitive Skills Navigation, (2) Software Engineering, and (3) Impact Testing. These processes must be executed cyclically and iteratively. This creates a global process that incrementally adjusts the technological tool that is being developed to the cognitive objective regarding the navigation of blind users.

Definition of Cognitive Skills Navigation: During this stage all of the O&M skills that are required to support the following are determined: perceptual development, spatial orientation, motricity, communication, basic concepts and protection techniques [17]. These skills are represented by behaviors and knowledge that the users must exhibit in order to carry out autonomous navigation.

This stage is fundamental to be able to begin the process for the development of technological tools. Here the most significant problems are identified, which are to be dealt with in the following software engineering process. The main objective of this stage is to determine the feasibility of the solution, as well as its restrictions (technical and methodological). The proposed solution will depend on the balance between the technological context and the orientation and mobility skills that are to be supported.

Impact Evaluation: We were interested in knowing the gains in terms of learning (O&M skills) between the pretest and posttest scores, which are the result of having used the application. The dependent variable corresponds to the O&M skills studied. Basically, this design responds to three steps: (1) Application of a pretest that measures the behavior of the dependent variable prior to the intervention; (2) Application of the intervention, which imply use of the video game for O&M skills; and (3) Application of a posttest that measures the behavior of the dependent variable after the intervention.

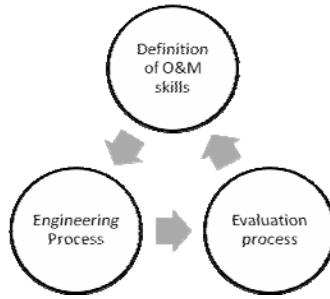


Fig. 1. Global iterative process for software development to improve the O&M skills in blind people

Software Engineering Process: In this stage of the model, efforts are mainly concentrated on the process of software engineering for the design and development of applications that can be used to improve the orientation and mobility skills of blind people. To these ends, it is proposed to work using the 5 traditional phases of systems development: Initial Phase, Analysis Phase, Design Phase, Implementation Phase and Evaluation Phase (see Fig. 2).

3.1 Initial Phase

During this phase, the feasibility for the development of the application is determined. An analysis of the technological context in which the video games will be developed is performed, and the O&M skills to be developed in users with visual disabilities are defined (see Fig. 2). **Technological Context:** Refers to the technology available in the market that can provide support for possible technological solutions for the development of orientation and mobility skills. In particular, the problem to be dealt with must be clear (defined by Orientation and Mobility Skills), and based on this one can define the most pertinent technology to be used. **O&M Skills:** The objective of this stage is to clearly define which specific O&M skills we will be able to support with the system to be developed.

3.2 Analysis Phase

In this phase an analysis of the end-users of the video game to be developed is analyzed, as well as the internal and external restrictions to the project. This is a stage to consider the most significant variables that will interfere in the design and development process. **Real Situations Analysis:** The real contexts in which the users would be able to carry out their activities, given the specific orientation and mobility skills that it is desired to support, must be analyzed. This analysis must include the problems that are presented within the environment, which could impede the navigational tasks from being completed, and considering a complete profile of the user that will utilize the system. **End-User:** The users' characteristics on a cognitive level (O&M), mental model, degree of vision and their most significant descriptive variables are specified. **Restrictions:** All of the restrictions that must be considered for the user to

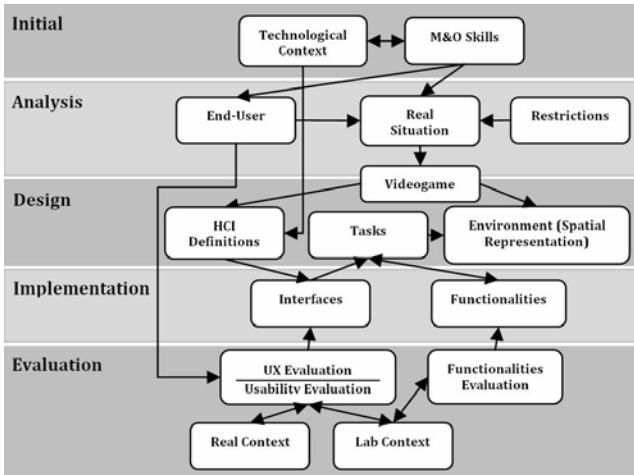


Fig. 2. Model to develop videogame applications to develop O&M skills in blind users

be able to correctly develop the desired O&M skills are defined. Both the rules of conduct for the user while using the technology and the social behaviors involved in the activity are specified. **Videogame:** This component, developed between the analysis stage and the design stage, allows us to define how to design the support system for O&M as a videogame.

3.3 Design Phase

HCI Definition: In this component the specific guidelines for how the interfaces and the interaction with the system to be developed must be designed. To do this, it is necessary to take the characteristics of the end users, their habitual ways of interacting and their interests into account. This also includes the specific characteristics of the system so that it has a recreational, “educational videogame” type orientation [12], and so that it is used to develop O&M skills. The end users’ considerations are provided during Analysis of the Real Situation stage, in which the users and their needs are defined (see Fig. 2). Due to this information, which can be considered within different technological contexts, it is necessary to include the way in which a blind user interacts with different kinds of technology [9]. **Environment:** Component responsible for defining where the system to be developed will be used. This is based on an abstract representation of the real world in the system. The computational representation of the real environment must be in accordance with the tasks that the videogame users have to perform. **Tasks:** For the correct development of the skills to be analyzed, it is necessary for the tasks, defined both on the level of the real environment (if the system is mobile, for example) and on the level of the software, to allow for the child to use the orientation and navigation tools through the videogame itself. This is essential for the kind of skills that it is sought to develop. These tasks must include elements from the real environment to be represented in the virtual world. General guidelines for how this environment should be represented will also be provided.

3.4 Implementation Phase

After the design phase the problem to be solved is clear, as well as the best way to achieve this. The task assigned to this phase consists of developing the solution that has been designed in the previous phase. **Interfaces:** In this stage of the model, the different interfaces that the blind user will use to develop the previously defined tasks are implemented. These interfaces can be of different sorts: Audio [14], haptic [8] and multimodal [4]. These interfaces must include the users' characteristics (from the Definition of HCI component), and the elements that make the orientation and mobility tasks possible. **Functionalities:** During this process, the data structures and the specific functionalities for the system being developed are defined. This component includes the activities within the tasks to be performed and implements the necessary functionalities, always making sure that it is possible to complete all the specific tasks assigned.

3.5 Evaluation Phase

By the end of the previous phase, the videogame has already been implemented. During this testing phase, the video game is tested in order to solve possible errors and defects (corrective maintenance), and to modify or improve the videogame (adaptive maintenance). During this stage, the following tasks must be considered. **Usability Testing:** In order to evaluate the interfaces used by the system developed, specific usability evaluations must be applied [9] (quantitative and/or qualitative), in order to assure that the users' interactions with the system are adequate and pertinent. These evaluations must be performed with end users, and involve the previously designed interfaces. **Real Context:** During the design of the interfaces for the system to be developed, it is necessary to consider evaluations based on the real environment. These evaluations can be either quantitative and/or qualitative. The main idea is that from these evaluations, relevant considerations regarding how the user interacts in the real environment emerge in such a way that the system adjusts to this kind of interaction. **Laboratory Context:** In order to make a more controlled and precise evaluation, it is necessary to perform experiments in a laboratory in order to evaluate usability. These evaluations would be of a more focused nature, and would define the specific redesigns for the interfaces being developed. **Functionalities Evaluation:** This stage of the development will validate if the functionalities of the system developed do what they have to do. Exhaustive tests must be taken within a laboratory setting regarding the system's behavior under various simulated conditions of use.

4 Conclusions

In this work, a model for the design, development and evaluation of video game-based applications is presented and described, so that users with visual disabilities can improve their O&M skills. A theoretical review of the concepts related to mobility and digital technology, the use of technology for O&M skills, and the use of video games to support learning was performed. Afterwards, the proposed model was presented, as well as its different stages and the impact that it has had on the development process.

Our previous experience with the design of educational software development models for the blind [39] has taught us how important it is to provide design and development tools for such systems. These tools can considerably improve the pertinence, acceptance and use of these systems by the end users.

The early development of orientation and mobility skills in blind children is fundamental for their performance in navigating unknown environments autonomously. At the same time, a higher understanding of space and the development of orientation and mobility skills does not only allow them to develop psychomotor activities at their age level, but also allows them to have a higher level of learning with regards to their perception and understanding of the environment. It is for this reason that the model presented in this work can be an essential contribution to generating video-games in support of the development of orientation and mobility skills, thus creating innovative ways of improving such abilities.

As it has been presented in the model, the end user takes on a leading role in the design of the systems, thus constituting an end user-based design. The participation of the end user in the main stages of development favors acceptance of the final product, an adequate and pertinent use in accordance with the user's mental model, and reduces the costs associated with redesigns due to problems with the interaction.

In future work, this model will be used as a basis for the design of mobile systems so that blind users can improve their navigational methods, making them more efficient and thus helping them to improve their autonomy in different, everyday-life environments. All in all, the systems based on this model will help to improve the navigation of blind users, and with this will aid in their integration and inclusion in society.

References

1. Amandine, A., Katz, B., Blum, A., Jacquemin, C., Denis, M.: A Study of Spatial Cognition in an Immersive Virtual Audio Environment: Comparing Blind and Blindfolded Individuals. In: Proc. of 11th ICAD, pp. 228–235 (2005)
2. Arnáiz, P.: Deficiencias visuales y psicomotricidad: teoría y práctica. Madrid: ONCE. Departamento de Servicios Sociales para Afiliados (1994)
3. Bradley, N., Dunlop, M.: Investigating design issues of context aware mobile guides for people with visual impairments. In: Proceedings of workshop on HCI in Mobile Guides Workshop at International Symposium, Mobile HCI 2004, Glasgow, UK, September 13, 2004, pp. 1–6 (2004)
4. Crossan, A., Brewster, S.: Multimodal Trajectory Playback for Teaching Shape Information and Trajectories to Visually Impaired Computer Users. *ACM Trans. Access. Comput* 1(2), 1–34 (2008)
5. Dowling, J., Maeder, A., Boldes, W.: A PDA based artificial human vision simulator. In: Proceedings of the WDIC 2005, APRS Workshop on Digital Image Computing, pp. 109–114, Griffith University (2005)
6. Lahav, O., Mioduser, D.: Construction of cognitive maps of unknown spaces using a multi-sensory virtual environment for people who are blind. *Computers in Human Behavior* 24(3), 1139–1155 (2008)
7. López, M.: Aspectos evolutivos y educativos de la deficiencia visual, 1a Edición. Netbiblo, SI, ACoruña (2004)

8. McGookin, D., Brewster, S., Jiang, W.: Investigating touchscreen accessibility for people with visual impairments. In: Proceedings of the 5th Nordic Conference on Human-Computer interaction: Building Bridges, NordiCHI 2008, Lund, Sweden, October 20-22, vol. 358, pp. 298–307. ACM, New York (2008)
9. Nielsen, J.: Usability engineering. Academic Press Professional, New York (1993)
10. Nielsen, L.: Oportunidades para el niño. Relaciones espaciales en niños ciegos congénitos. In: International Council for Education of the Visually Handicapped, vol. 66, pp. 3–15. Christoffel Blinden Mission (1989)
11. Ochoa, S., Alarcón, R., Guerrero, L.A.: Understanding the Relationship between Requirements and Context Elements in Mobile Collaboration. In: Jacko, J.A. (ed.) HCII 2009. LNCS, vol. 5612, pp. 67–76. Springer, Heidelberg (2009)
12. Prensky, M.: Computer games and learning: Digital game-based learning. In: Raessens, J., Goldstein Raessens, J. (eds.) Handbook of Computer Game Studies, pp. 97–122. The MIT Press, Cambridge (2005)
13. Rodrigues, C.: Um Dispositivo Háptico de Auxílio À Navegacao para Deficientes Visuais. Trabalho de Graduacao ao Centro de Infomática da Universidade Federal de Pernambuco para a obtencao do grau de Bacharel em Ciência da Computacao (2006)
14. Sánchez, J.: User-Centered Technologies for Blind Children. Human Technology Journal 45(2), 96–122 (2008)
15. Sánchez, J., Maureira, E.: Subway Mobility Assistance Tools for Blind Users. In: Stephanidis, C., Pieper, M. (eds.) ERCIM Ws UI4ALL 2006. LNCS, vol. 4397, pp. 386–404. Springer, Heidelberg (2007)
16. Sánchez, J., Sáenz, M., Baloian, N.: Mobile Application Model for the Blind. In: Stephanidis, C. (ed.) HCI 2007. LNCS, vol. 4554, pp. 527–536. Springer, Heidelberg (2007)
17. Schultz, P.: Movilidad e independencia para el discapacitado visual. Dinámicas Psicológicas del proceso de Enseñanza. Córdoba (Argentina): ICEVH. vol. 69 (1990)

Underestimated Cerebral Visual Impairment Requires Innovative Thinking When Using AT

Michael Cyrus and Frank Lunde

Senior Advisors

NAV National Center for Assistive Technology, Oslo, Norway
{michael.cyrus, frank.lunde}@nav.no

Abstract. The impact of cerebral visual impairment (CVI) obviously is widely underestimated and knowledge about CVI is not widespread. This article illustrates in which way the dysfunction of visually guided handmovements as well as simultaneous attention demands rethinking hardware design especially of input devices. CVI dysfunctions are manifold and special software is needed. The challenges of some aspects of CVI may be met by everyday hardware as modern smartphones providing numerous software which might show to be useful, even if designed for other purposes than CVI-rehabilitation. It is possible that useful software is already created (or may be created) when the open system invites to creative program development from the public. The role of user centered design should therefore be given more attention.

Keywords: CVI (cerebral visual impairment), visually guided hand movements, simultaneous attention, writing difficulties, user centered design, alternative input devices, iPod.

1 Usable User Interface for Persons with Cerebral Visual Impairment

Cerebral visual impairment (CVI) is a widespread but little known condition with various facets and degrees. New technology in everyday life and work creates challenges to our abilities to visual sensory organising and understanding. Modern man interacts with Automatic Teller Machines or Self service web pages meant to rationalize processes which used to involve contact and communication between people. All these new interfaces challenge visual processes as well as visually guided action and movement. Computers may be obstacles as well as aids to overcome challenging situations.

In our country, knowledge about CVI and systematic rehabilitation for CVI problems is limited and unknown among designers. Wrong diagnoses may probably be common. It may well happen that individuals are identified with cognitive problems, motor disabilities and other forms for impairment rather than having the central key condition identified. An individual may function and in certain ways act as a blind person even though he/she obviously have functional vision left. There may be specific situations where vision does not work as it is expected to do. Defects may be found in

the ability to perceive or naming specific visual features, recognizing specific features of objects, recognizing people, finding one object among others, ranking, visual memory, visually guided movements, visual organization and visual construction.

2 Rehabilitation of CVI Still Is Pioneer Work

In later years, the knowledge of cerebral processes has increased considerably due to neuroimaging. Most of the work done in the field of CVI is connected to the study of prevalence of certain types of CVI, as well as methods of measurement and identification. Not much has been done within the field of rehabilitation of individuals afflicted with this condition. Although CVI has become more widely discussed throughout the last 15 years, still work connected to universal design, adaptation of workplaces or learning processes in adolescents and adults are so to speak nonexistent. However, some early work is done by Groenfeld and Jan in Vancouver, Jacobson at Karolinska in Stockholm and in Northern Italy Fazzi and Ugetti, and G Dutton in Scotland. More recently in Norway by Wilhemsen on persons with impaired eye-movement and impaired reading ability after stroke. One of the authors of this paper has done some rehabilitation work with children with CVI caused by PVL at the dep. of Child neurology at Rikshospitalet University Hospital and later at his present workplace (not published). In Scotland work is done at centers in Glasgow and Edinburgh (Scottish Sensory Center - <http://www.ssc.education.ed.ac.uk>) Software and hardware design may be a part of the rehabilitation work - at least the universal design thinking.

3 An Illustrating Case

The case shows the misunderstandings and pitfalls CVI may provide for the rehabilitation system. It shows that platforms like iPhone/iPod and Android with their possibilities for everyone to create software may push rehabilitation-technology toward a more democratic level of development than ever before - not depending on initiatives of big companies but to individuals with the interest to create software even individually tailored for people with special needs.

4 Medical Conditions

Our client is a 15 years old boy with cognitive abilities adequate to his age. Born in term. Our client's GP contacted SIKTE when Our client's family and helpers «had tried everything» and he was still not able to write. Reading through all assessments of our client we found following results:

- Early observed trouble with hitting things he wanted to grip or do finemotor activities. He could not hit the correct keys on a keyboard. The observed handmotor problems showed no correlat under MR of motorcortex. (That means, he has not a real handmotoric problem.)
- Slow on movement and execution of motortasks.
- Problems with balance and coordination tasks.

- Troubles with antagonistic movements, ataxia when he comes close to the aim of his movement. Problems with simultaneous attention between several objects and processing auditory and visual information.

Our client's main problem was that he was assessed as a person with motor impairment, while in fact he had an impairment connected to visual processing and deficits in visuo-cognition. In his case it was a problem called optic ataxia, which means that the individual meets severe challenges to get his hands where he wants to have them while steering them visually guided towards the object. The hands' movement systematically will end near the object, but not touch it.

5 The Solution

Our client is interested in technology, and has always shown great interest in experimenting with high-tech devices and finding out how they work. He got a new mobile phone from his mother and experimenting with it he found using the phone in landscape mode (full QWERTY) that he could write sentences which were close to correct orthography and understandable. This observation – combined with doctor's notes – became the key to understand his writing problem and the solution to it.

This ability was important to the solution that finally made him able to input text into the computer. We could not find any way of using his phone as an input device for the computer.

An iPhone app developed by Evan Schoenberg MD did the job. It is called Rowmote Pro and may be downloaded from the App Store. The App works only with an Apple Computer with OSX ver 10.5 or later. The iPhone communicates with the computer through WiFi and inputs mouse operations, remote control operation and keyboard to the computer. The computer has to be connected to the same network as the iPhone. When writing, the computer will read the words using VoiceOver that comes with OSX ver. 10.5 and later. This made it possible for him not to look at the computer screen while writing.

Our client operates the keyboard using his thumbs. He is holding the phone with both hands, and moves only his thumbs when writing. He is sliding the thumb over the keyboard. Selecting the letter is done when lifting the thumb from the keyboard's position of the letter he is writing. He uses his memory for movements he made with his thumb to find the actual letter to write. As he was not so thoroughly trained he sometimes has to use vision to find the keys, but not able to steer the fingers towards them with visually guided movements, so-called optic ataxia.

Problems with simultaneous attention between several objects as well as slow motor execution gave him challenges shifting gaze between the keyboard – as little as it is – and the field where text occurs. It may therefore be an advantage to get auditory feedback from keypositions like we know it from screenreaders for visual disabled persons. The iPhone 3Gs has this possibility. It would then be possible to hear the name of the key over which the one thumb is hovering, and the other thumb then affirms by tapping somewhere on the screen. Experienced blind users describe this way of writing as very effective. Unfortunately we only had an older iPhone version which did not have this possibility, but we are sure that he would have taken advantage of this solution,

and – writing so to say in the blind – he would be able to concentrate on the text on the screen. We believe that given some time he will be more efficient when writing.

6 Discussion and Conclusion

Even under good conditions, the larger keyboard did not work. Our Client could not figure out how to use the the large numbers of keys – even when they were enlarged. This again was the result of his problem with visual guidance of hand movements. The problem with paying simultaneous attention towards several objects might also have had an impact.

Another suggested solution was to use a desc-top-screen with an on screen-keyboard. This worked also under the condition, that his hand was supported by a cushion or a similar aid. The disadvantage was the immobility of this set, which is impractical for a student who has to migrate between rooms during the school-day. Optic ataxia and problems with simultaneous visual processing and visually guided movements are aspects of CVI. The remedy meant to be a helper turned out to be the obstacle just by the same means by which it should help. Alternative design of input devices was a key to the final solution. Probably this case also might be relevant for thinking universal design as simultaneous attention, as well as other specific dysfunctions might be more widespread as usually recognized.

But it is not only the knowledge on CVI which turns the situation to a happy end. It is also the possibility for anyone to make software on mobile platforms, which inspires creative persons to make useful applications, and their work is easily available through the channels provided by the manufacturer of the devices. This opens new perspectives on computers helping people: Applying relative easily designable and economically and physically available software for multiple tasks in mobile and personal software.

References

1. Dutton, G.N., Jacobson, L.K.: Cerebral visual impairment in children. *Semin. Neonatol.* 6(6), 477–485 (2001)
2. Good, W.V., Jan, J.E., DeSA, L., Barkovich, A.J., Groenfeld, M., Hoyt, G.S.: Cortical vision impairment in children. *Survey of Ophthalmology* 38(4), 351–364 (1994)
3. Groenfeld, M.: Children with cortical visual impairment. Materials distributed at cortical visual impairment: Assessment and implications for education, pp. 1–23. Royal Institute for Deaf and Blind Children, Sydney (1997)
4. Groenfeld, M., Jan, J.E., Leader, P.: Observations on the habilitation of children with cortical visual impairment. *Journal of Visual Impairment and Blindness* 84(1), 11–15 (1990)
5. Liu, G.T.a.o.: *Neuro-Ophthalmology. Diagnosis and Management*, Saunders (2001)
6. Wilhelmsen, G.B.: *Å se er ikke nok*, Unipubforlag, Oslo (2003)
7. Application Rowmote - Apple app store

Eye Tracking for Low Vision Aids - Toward Guiding of Gaze

Yasuyuki Murai¹, Masaji Kawahara², Hisayuki Tatsumi^{2,*},
Iwao Sekita², and Masahiro Miyakawa²

¹ Nihon Pharmaceutical University
10281 Komuro, Ina-cho, Saitama, Japan 362-0806
murai@nichiyaku.ac.jp

² Tsukuba University of Technology
4-12-7 Kasuga, Tsukuba, Ibaraki, Japan 305-8521
{masaji, tatsumi, sekita, mamiyaka}@cs.k.tsukuba-tech.ac.jp

Abstract. Eye tracking technique in the visibility study of low vision was newly introduced in the previous report, where we examined the ease of finding public signs on the streets and in the interior of buildings by low vision people. We got a conclusion that they hardly notice these signs. In this report we continue our research in this direction. We describe details of eye tracking technology applied to low vision. We devise calibration method for low vision. We describe analysis of eye tracking data on the basis of simplified gaze circle model of sight of low vision, leading to a conclusion that it is possible as well for low vision to locate regions of interest (ROI) by applying classical method of scanpath analysis. We also show a preliminary result of public sign recognition in the view by using a fast pattern matching technology called “boosting,” linking to a future system for guiding the gaze of low vision to a missing public sign and zooming into it.

Keywords: Low vision, Eye tracking, Visual aids, Public signs, View image segmenting.

1 Introduction

Barrier-freeing is going on steadily in Japan, especially since after so-called “New barrier-free law” in 2006. However, from sighting standpoint, usability of public places has not yet been barrier-freed for everyone. Public signs contribute people for accessibility. In the previous report [7], using eye tracking system worn by a low vision subject we examined ease of finding out public signs on the streets and in the interior of public buildings. We found that low vision people are hardly noticing the public signs while walking even under clear weather condition.

We focus our attention on recognizing public signs by low vision people. Eye tracking is one of the latest technologies that has shown potential in several areas including

* This research was partially supported by the Japanese Ministry of Education, Science, Sports and Culture, Grant-in-Aid for Scientific Research (B) 21300079, 2009; (C) 20500798, 2008.

human-computer interaction for people with and without disabilities, and for noninvasive monitoring, detection, and even diagnosis of physiological and neurological problems in individuals[6]. Previous studies using eye tracking mainly concern with the role of the eye in reading, and so they are oriented only to the sighted [5].

In this report, we describe details of eye tracking for low vision people. Eye tracking is not available for all of them (for example, for nystagmus no eye tracking is possible). We show calibration method of eye tracking apparatus for low vision. We model visibility of low vision simply by gaze circle (instead of gaze point). Then, taking our low vision model into consideration, we analyze distribution of gaze duration time on the scanpath, leading to the conclusion that it is possible to localize regions of interest (ROI) by this method. Finally, we show our preliminary result of guiding gaze of low vision to a missing for him public sign by automatically detecting it in the scene. The result seems to suggest possible application of eye tracking technology[2,5] for low vision aids, for automatically detecting and segmenting out navigation and guiding clues in the scene.

2 Preliminaries

2.1 Public Signs

Public signs are set up for the convenience of people. Often pictograms (or pictographs) are used instead of letters to catch the attention of the public. We show an example pictogram in Fig.1. They work for navigation, guiding, explanation, regulation and emergency purposes. So recognition of them is sometimes crucial, especially for the visually impaired people. We focus our attention on the recognition of public signs.



Fig. 1. Pictogram

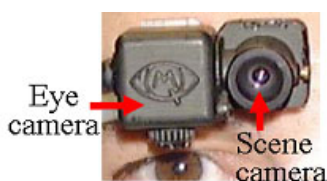


Fig. 2. Optical unit (two cameras)

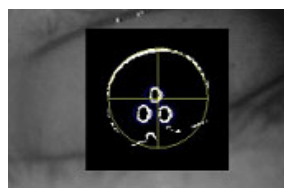


Fig. 3. Calibration of pupil

2.2 Eye Tracking System

This is a system that can record moves of gaze point of a human being with short intervals automatically. In our experiment we have used Mobile Eye (a product from ASL Co.[1]), which consists of two components: Optical unit and Recording unit. The former, being attached to the spectacle (shown in detail in Fig.2) consists of two cameras: one is the *scene-camera* which records the environment being observed by the user while the other is the *eye-camera* which records the eye being tracked. Both image streams are stored in the recording unit (on the same videotape by alternating frames at 60Hz). So the actual sampling is 30Hz. Note that these eye tracking apparatuses could become of the size of glasses and useful tools for the low vision aids in future.

3 Eye Tracking for Low Vision

To get environmental information we rely largely on visual information. This is also true for low vision people: if slightest residual sight is applicable, people persistently rely on it. So assisting low vision people is most effective if we do by use of their residual vision.

For improving their sight monocular or low-vision glasses are used. These optical apparatuses enlarge visual field, so the gaze point of the user moves large distance in the vision quickly. This causes unpleasant feeling and the user becomes fatigue quickly. Target could be recognized by him if it is caught within the scope of his sight, but tracking the target needs enormous efforts. On the basis of these observations, we have set the following three conditions on the assisting low vision technology in our project: it should help a low vision user 1) to focus gaze point or gaze-direction, 2) not only with sight but also with visual field, 3) with the task of finding a target (e.g. public signs) while walking on the streets. Eye tracking technology for low vision people attracted our attention, since it utilizes existing residual sight of low vision user maximally and at the same time it fulfils our three conditions, especially 3).

3.1 Calibration for Low Vision

Eye tracking is obtained using a pupil-corneal-reflection system. The optical unit of the Mobile Eye consists of two cameras: eye-camera and scene-camera Fig.2. The latter observes the view of the user, while the former observes the pupil of the user, which serves for the calculation of the gaze point direction. Infrared ray emitter attached near eye camera emits rays into the pupil of the user, and their reflection illuminates on the pupil. Eye camera detects and measures moves of the pupil by observing the illumination Fig.calibration-of-pupil. By synchronization of the two optical images, gaze point is placed on the scene correctly with a sufficient accuracy. The synchronization is done through a calibration of the human-optical system before the experiment. It is done by the following procedure: first pupil of the user is recognized by the eye camera. Then several points (usually five to ten, and at least four) are distributed in the view of the view camera. User is requested to gaze each calibration point. At the same time the pupil is observed by the camera. On the basis of these data and the direction of pupil measured by the eye camera, the location of gaze is decided on the scene (the crossing point of the cursor lines indicates the gaze point in Fig. 4(a)).

Sometimes it is difficult to calibrate the vision of a low vision user because the pupil-analysis does not work well. We have tried calibration of the eye tracking system for severe low vision users whose corrected visions are less than 0.1 (in total 4 persons). For those with nystagmus (involuntary eye movement) the calibration was not operable. However, for those without nystagmus we had a conclusion that it is possible to do the calibration, though we have to sacrifice precision to a certain degree. The major difficulty with low vision is to locate gaze point precisely on the designated point. In Fig.4(b) we show an example of charts which we have used for the calibration. A low vision user is asked to read characters in the chart, which is projected on the screen placed in front of him. The characters in the chart are magnified until he can read them. The gaze point of the user we assumed at the center of the character. For his readability

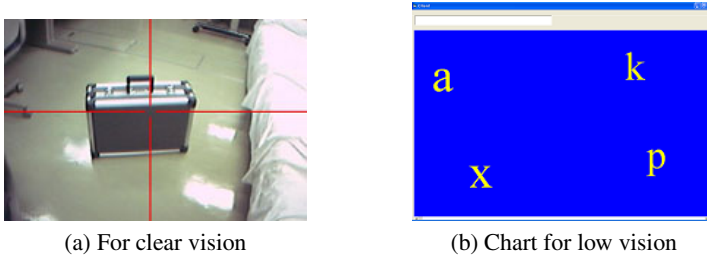


Fig. 4. Gaze point calibrations for eye tracking system

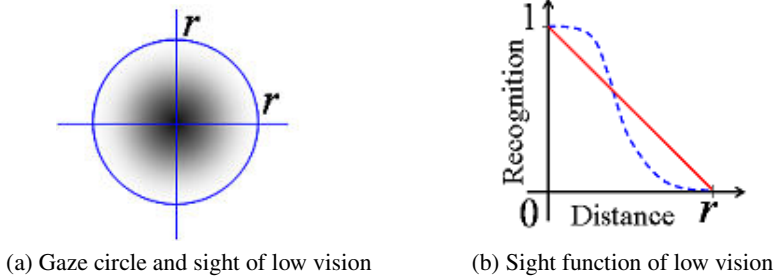


Fig. 5. Gaze circle model of the sight of low vision

the characters in the chart must be simple alphabets or numerals, with colors changed or with black-white inversion. For more precision, we could use Landolt rings. To cope with this difficulty we are asking ASL to loosen the definition of gaze point used in the calibration from point to small region (area) with varying size so that we can accommodate low vision users in the calibration procedure.

3.2 Gaze Circle Model for Low Vision Sight

Now we model the sight of low vision for eye tracking analysis. Instead of a gaze point, we assume a gaze circle with a small radius r for low vision, where r is a parameter determined by the sight of a low vision person (Fig. 5(a)). We assume that the sensitivity of sight of low vision person decreases inversely proportional to r , i.e. we assume that it is represented by

$$1 - d/r, \text{ where } d \in [0, r]. \tag{1}$$

It may be more adequate to approximate this function by a pseudo-logistic function depicted by dotted-line in Fig.5(b). However, this function is too complicated to meet our application shown later.

4 Analysis of Eye Tracking of Low Vision

One of the objects of our project is to develop a system that zooms in an object that is hard for a low vision person to recognize, yet it is likely that his sight has caught it in

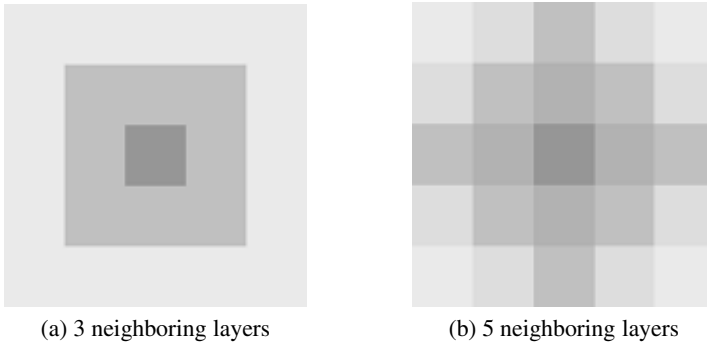


Fig. 6. Two approximations for distributing duration of gaze to the neighbors

the field of view. To achieve this, we need to detect an object which he is searching for. A typical behavior of his sight during the search is intensive duration time of his gaze around a small region in the view.

In the analysis of eye tracking, the traces of the moves of gaze point is called *scan-path*. We show an example of scanpath for low vision in Fig.7(a). In the scanpath, we want to locate several regions in which he is interested or searching for something. Such regions are called *regions of interest* or ROI[5].

We have developed a program which plots duration time of gaze on each small region centered by a gaze point. We took 5×5 pixels as a unit region. We accumulate duration time of gaze for each region. In the counting process, taking the sight of low vision people approximated in the previous section into account, we have distributed the duration time on a gaze point over its neighboring regions as well. As the distribution weights (rates) we might have taken values determined by the sight function expressed by the formula (1). However, this is not yet done. Instead, we have taken two variations shown in Fig.6 as neighbors, i.e. (a) three-layer model and (b) five layer-model. In both models, no distribution is given to the regions outside the outer-most layer (to the 4-th or further layers or to the 6th or further layers). Which of the variations of distribution is to be adopted is determined depending on r , the sensitivity of the sight of the user.

The result of the counting is drawn on the scene in 25 grades. In Fig.7(b) we show the resulted distribution of duration time of gaze for the view in Fig.7(a) (view of the reception of the library) . The darker the color of region, the longer the staying duration time of gaze on the region. Subjects are told to keep their heads fixed during the time of searching objects. In most cases we could have a good scanpath, as low vision people take longer time of viewing than normally-sighted people. As a conclusion we can determine regions of interest (ROI) for low vision users by using the eye tracking technology.

We are preparing for the segmentation of regions of interest (ROI) and zooming into it. They can be displayed on the mobile monitor, so that user can recognize it by his sight on site (verbal notification of the public signs and reading up of the characters using the character recognition may be helpful). Various picture enhancement technologies could be applied in order to help those people with other irregularities in their sight.

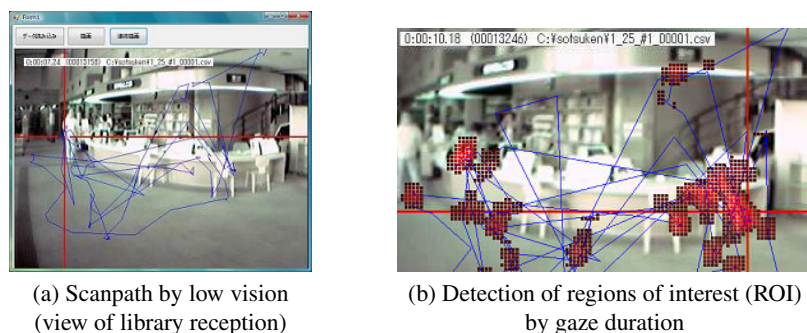


Fig. 7. Scanpath analysis for low vision and detection of ROI by gaze-duration

5 Toward Guiding of Gaze of Low Vision

We have asked a question to our students, the users: What would you do if you can't recognize an object which you are seeking in the vision? One of the answers was the following: If he could take a picture of the scene by a camera attached to the mobile phone, he could see and recognize it by moving his eyes close to the monitor display. However, in practice, it is hard for a low vision person to focus his camera on the object, even though his gaze has certainly been turned to the direction of the object.

Guiding of gaze to object or to object direction is a powerful assistance to those people whose impairment accompanies narrowed or deficit field of vision. For a person with profoundly impaired field of vision, it is an extremely difficult task to catch object in his vision. So, instead him, let assisting devices do this task for him.

Assume that our objective is to find out pictograms. Then for each pictogram registered in the database we must do pattern matching with the picture of the view. This is not a simple task; there arise several problems.

Firstly, each pictogram is not standardized. There are several pictograms for "toilet"; often "for men" is indicated with different designs, and the same is true "for women" (we need to register all pictograms as correct patterns which appear along the pathway of the user). Secondly, it is a more difficult task for searching algorithms to end the processing by judging that there is none the target pictograms in the view rather than finding them. Thirdly, the process of pattern matching may require enormous time if we could not get a stationary (still) field of view. A more high sampling rate of eye tracking picture may be required obeying the request of the assisting devices (current apparatus works at the rate of 30 Hz).

In short, it is likely that perfect solution is still beyond the reach considering the state of the art of pattern recognition technology. Toward practical solution of achieving a step-by-step improvement of recognition, we adopt the following strategies: 1) We apply general-purpose algorithms rather than various case-bounded special ones, 2) We are expecting that we can eliminate erroneous judgements step-by-step by incorporating learning of patterns into our algorithms.

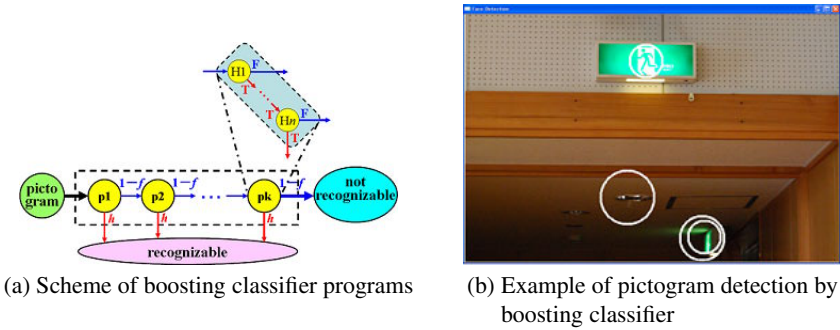


Fig. 8. Towards guide of gaze

Recently there is a progress in object matching area. A technology is extensively used for focusing human faces in digital camera applications. In this technology a learning method called “boosting” is used [3], where many small procedures with small detecting powers are combined in cascade into a detector with a large detecting power (Fig.8(a)). In the learning stage of each small classifier, special features of patterns called Haar-like features are used either in horizontally scaled forms or in vertically scaled forms. A program library is operating for free use in Open CV (Open Source Computer Vision Library). Using this library one can build face-detection or object detection programs with small elaboration effort [8].

In Fig.8(b) we show a sample result of detecting an emergency exit pictogram in the picture of gymnastic hall (the detected objects are shown in circles). The detection program is applied to the sample picture after it has learned about the features of a correct sample target pictogram for 1 - 2 hours. Possibly because of insufficient learning, still other regions are also detected in this picture as the target object. We think that it should be easy to remove these incorrect candidates using other useful features of the target object.

6 Conclusions

The purpose of our project is to develop assisting technologies for the sight of low vision people. Motivated from our previous result that public signs are hardly recognized by low vision people[7], we have investigated eye tracking technology for applying public signs recognition purpose by low vision people. We have shown that eye tracking technology is applicable to low vision people as well as to the sighted and has potential to be used for helping their sight by locating regions of interest (ROI) in the view. We have also shown that our preliminary program using learning and boosting pattern classifier successfully segments out candidates of the object pattern from the scene. This suggests a potential use of eye tracking technology for the navigation of low vision walking.

Our next task includes enhancing and enlarging the pattern of signs and sending it back to the user’s mobile monitor for his own recognition by his sight.

References

1. Applied Science Laboratories, product: Mobile Eye, <http://www.asleyetracking.com/products/mobileeye.htm>
2. Paulson, E.J., Goodman, K.S.: Influential studies in eye-movement research (January 1999), <http://www.readingonline.org/research/eyemove.html>
3. Viola, P., Jones, M.: Rapid object detection using a boosted cascade of simple features. In: Proc. 2001 IEEE Conf. on Computer Vision and Pattern Recognition (CVPR 2001), vol. 1, pp. 511–518 (2001)
4. Lienhart, R., Maydt, J.: An extended set of Haar-like features for rapid object detection. In: Proc. 2002 IEEE Int. Conf. on Image Processing (ICIP 2002), vol. 1, pp. 900–903 (2002)
5. Jacob, R.J.K., Karn, K.S.: Eye tracking in human-computer interaction and usability research. In: Hyona, J., Radach, R., Deubel, H. (eds.) *The Mind's Eye: Cognitive and Applied Aspects of Eye Movement Research*, pp. 573–605. Elsevier Science, Amsterdam (2003)
6. Talukder, A., Morookian, J.-M., Monacos, S., Lam, R., Lebow, C., Bond, A.: Real-time non-invasive eye tracking and gaze-point determination for human-computer interaction and biomedicine. In: SPIE Defense and Security Symposium, Optical Patter Recognition XV, Orlando, FL, April 12-16 (2004), <http://hdl.handle.net/2014/38027>
7. Tatsumi, H., Murai, Y., Sekita, I., Miyakawa, M.: Public signs sight assessment for low vision through eyetracking. In: Miesenberger, K., Klaus, J., Zagler, W.L., Karshmer, A.I. (eds.) *ICCHP 2008*. LNCS, vol. 5105, pp. 138–141. Springer, Heidelberg (2008)
8. NAIST (Nara Institute Science and Technology) OpenCV programming book team: *OpenCV Programming Book for OpenCV 1.1*, 2nd edn., Mainichi Communications (2009)

Assistive Technologies as Effective Mediators in Interpersonal Social Interactions for Persons with Visual Disability

Sreekar Krishna and Sethuraman Panchanathan

Center for Cognitive Ubiquitous Computing (CubiC), Fulton School of Engineering,
Arizona State University, Tempe AZ, USA
{Seekar.Krishna, Panch}@asu.edu

Abstract. In this paper, we discuss the use of assistive technologies for enriching the social interactions of people who are blind and visually impaired with their sighted counterparts. Specifically, we describe and demonstrate two experiments with the Social Interaction Assistant for, a) providing rehabilitative feedback for reducing stereotypic body mannerisms which are known to impede social interactions, and b) provide an assistive technology for accessing facial expressions of interaction partners. We highlight the importance of these two problems in everyday social interactions of the visually disabled community. We propose novel use of wearable computing technologies (both sensing and actuating technologies) for augmenting sensory deficiencies of the user population, while ensuring that their cognitive faculties are not compromised in any manner. Computer vision, motion sensing and haptic technologies are combined in the proposed platform towards enhancing social interactions of the targeted user population.

Keywords: Assistive Technology, Social Interactions, Dyadic Interpersonal Interaction, Computer Vision, Haptic Technology, Motion Sensors.

1 Introduction

Social interactions are part and parcel of everyday living for both personal and professional success of one's life. From a neurological perspective, social interactions result from the complex interplay of cognition, action and perception tasks within the human brain. For example, the simple act of shaking hands involves interactions of sensory, motor and cognitive events. Two individuals who engage in the act of shaking hands have to first make eye contact, exchange emotional desire to interact (this usually happens through a complex set of face and body gestures, such as smile and increased upper body movements), determine the exact distance between themselves, move appropriately towards each other maintaining Proxemics (interpersonal distance) that are befitting of their cultural setting, engage in shaking hands, and finally, move apart assuming a conversational distance which is invariably wider than the hand shake distance. Verbal exchanges may occur before, during or after the hand shake itself. This example shows the need for sensory (visual senses of face and

bodily actions, auditory verbal exchange etc.), perceptual (understanding expressions, distance between individuals etc.), and cognitive (recognizing the desire to interact, engaging in verbal communication etc.) exchange during social interactions.

Individuals who are disabled face myriad levels of difficulty during everyday social interactions depending on the kind of disability. This is due to the fact that nearly 65% of all human interpersonal communications happen through non-verbal cues [1]. Non-verbal cues are mostly interpretative and not instructive as verbal cues (such as speech) are. In a bilateral interpersonal interaction, while speech encodes all the information, non-verbal cues facilitate an elegant means for delivery, interpretation and exchange of this verbal information. People with sensory, perceptive, motor and cognitive disabilities may not be able to receive or process these non-verbal cues effectively. Though most individuals learn to make accommodations for the lack of a primary information channel, and lead a healthy personal and professional life, the path towards learning effective accommodations could be positively effected through the use of assistive aids. In this paper, we focus on the topic of building an assistive aid for social interactions and discuss how they could enrich interpersonal communications. We specifically focus on the issues emanating from the lack of sensory visual channel, like in the case of people who are blind or visually impaired, and describe the design and development of a social interaction assistant that conveys important dyadic non-verbal communication cues.

2 Related Work

The need for developing social interaction assistive technologies was the fall out of two focus group studies that we carried out with individuals who are blind and their abettors [2]. People who were blind identified social situations where visual disability stymied their ability to reach out to their sighted counterparts. Similar emphasis on the need for social interaction assistance was reported by Shinohara et al. [3], where, the researchers followed a college student who is blind and categorized the important needs for this individual in her day-to-day life. Jindal-Snape et al. [4] has highlighted the need for social assistance for children who are blind and visually impaired. Their work in India has tied the lack of sensory input to critical factors that lead to the development of socially distracting behaviors, like body rocking. These unfavorable behaviors are precursors for social isolation and related psychological problems. To the best of the authors' knowledge, only one related work focuses on the issue of delivering social information to people who are blind: In [5], Réhman et al., developed a haptic chair for presenting facial expression information to people who are blind. A chair was equipped with vibrotactile actuators on its back rest forming an inverted Y. A camera mounted elsewhere tracks the mouth of an interaction partner, and actuators vibrate along any one of the three axes of the Y, based on whether the interaction partner was neutral, happy, sad or surprised. No formal experiments were conducted with the target population. Further, this solution had the obvious limitation that users need to be sitting in the chair to use the system. In this paper, we illustrate how social interactions have myriad levels of non-verbal cueing, where facial expressions play a significant yet incomplete role. We discuss a social assistive technology framework and elaborate on two solutions that can enrich social interactions for the targeted user community.

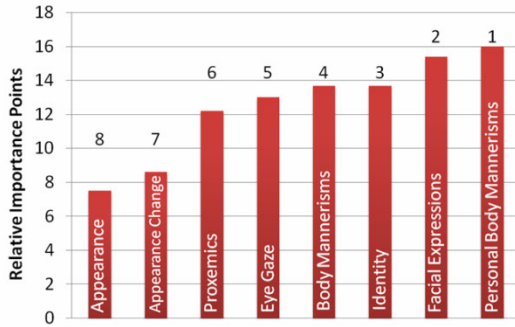


Fig. 1. Self report importance (scaled over 100 points) of visual non-verbal cues obtained through an online survey of target population and specialists [6]



Fig. 2. Embodied Social Interaction Assistant Platform

3 The Social Interaction Assistant

In [6], we identified 8 important social needs for individuals who are blind and visually impaired and rank ordered them through an online survey. A potential assistive technology platform for enriching social interactions was then proposed in [7], which relies on state-of-the-art wearable camera technology and computer vision algorithms for extracting important social interaction cues. In Fig. 1, we show the rank ordered

needs list as graded by a group of individuals with visual impairment. This list shows that participants' most important need corresponds to feedback on their own body mannerism and how it was affecting their social interactions. Following this was their need to access facial expressions, identity, body mannerisms, eye gaze, proxemics and appearance of their social interaction partners, in the presented order.

Fig. 2 shows the integrated social interaction assistant platform that we have developed towards enriching the social interactions of people who are visually impaired. A camera mounted on the nose-bridge of a pair of glasses, a wireless motion sensor mounted within the clothing of the users, and a handheld button based user input device act as inputs into the system. Any processed social scene information is either delivered through a pair of headphones or through vibrotactile actuation devices including a haptic belt [8] and a haptic glove [11]. In this paper, we discuss two important components of the prototype, corresponding to rank ordered needs 1 and 2, from the above chart. We have discussed solutions for need 3 in [8], while solutions for need 6 are addressed in [9].

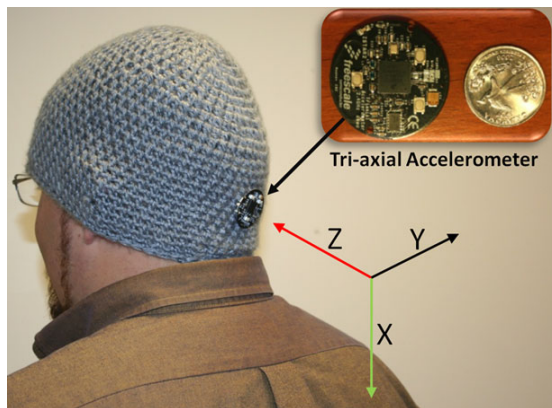


Fig. 3. Motion sensor for detecting body rocking stereotypy

3.1 Components of the Social Interaction Assistant and Associated Research Questions

3.1.1 Detection of Stereotypic Body Mannerisms

Corresponding to the need 1 depicted in Fig. 1, in order to detect user's non-social body mannerisms, we resort to the use of clothing-integrated motion sensors. Eichel [10] introduced taxonomy of stereotypic body mannerisms that people with blindness and visual impairment tend to display and identified that body rocking as one of the most commonly seen behavior stereotype. In order to detect body rocking, as shown in Fig. 3, we incorporated a wireless accelerometer on the social interaction assistant to detect and categorize upper body movements. Through the use of supervised pattern recognition algorithms, based on adaptive boosting of triaxial acceleration data, rocking pattern is isolated from other functional body movements, including bending, stooping, leaning etc., which involve motion similar to that of rocking but are functionally relevant.

Research Question: How well can body rocking be distinguished from other functional movements by using on-body motion sensors?

3.1.2 Conveying Facial Expressions of Interaction Partners

Following the need to realize their own body behaviors, individuals surveyed appealed for the need to understand the facial mannerisms and expressions of their interaction partner. Extracting and delivering facial mannerisms is a complex computational/engineering task. The human face is very dynamic when it comes to generating non-verbal communicative cues. Subtle movements in the facial features can convey great amounts of information. For example, slight opening of the eyelids conveys confusion or interest, whereas a slight closing of the eye lids conveys anger or doubt. In the past decade, computer vision research has achieved encouraging levels of capability in recognizing some of the human facial expressions. Little work (Ref. [5]) has been done towards finding a means of communicating this information back to people who are visually impaired. Most researchers and technologists have resorted to auditory cueing; but there is a strong growing discomfort in the target population when it comes to overloading their hearing. In the proposed approach, we explore the use of vibrotactile cueing (haptic/touch based cueing) on the back of the human palm (the human hand has a very large representation in the somatosensory cortex of the brain) to deliver any facial expression information. Camera on the wearable glasses (Fig. 2) is used for extracting the seven basic expressions (Happy, Sad, Surprise, Anger, Neutral, Fear and Disgust) of an interaction partner, which are then encoded as haptic cues. Fig. 4 shows the vibrotactile glove used in the social interaction assistant and the corresponding mapping of seven basic expressions. Please refer to [11] for details.

Research Question: How well can the seven basic human expressions be conveyed through haptic icons?

4 Experiments and Results

4.1 Detection of Body Rocking

Rocking motion patterns and functional movement (like bending and stooping) patterns were collected on 10 participants for duration of 5 minutes each. A pattern recognition engine based on AdaBoost learning algorithm was trained to distinguish between rocking and non-rocking patterns. All experiments were conducted on time slices of the motion data by splitting the motion data into discrete packets of samples.

Fig. 5 shows the overall detection rates as a function of time slice length in seconds. It can be seen that the best performance of our device was to detect rocking from non-rocking on an average of 95% with a latency of 2s. It was possible to detect rocking at a much less latency (0.5s), but the detection rate went down to 92%. Thus, 95% of the time, we could inform a subject of his/her rocking activity within 2 seconds of them starting to rock. Further, we noticed that the average natural rocking motion of all 10 subjects was 2.22 seconds per rock, which implies that a latency of 2 seconds was well within the time duration of a single rocking action.

While tests were done on rocking as a stereotype, any bodily movement can be modeled similarly by choosing an appropriate location of the motion sensors. Similar

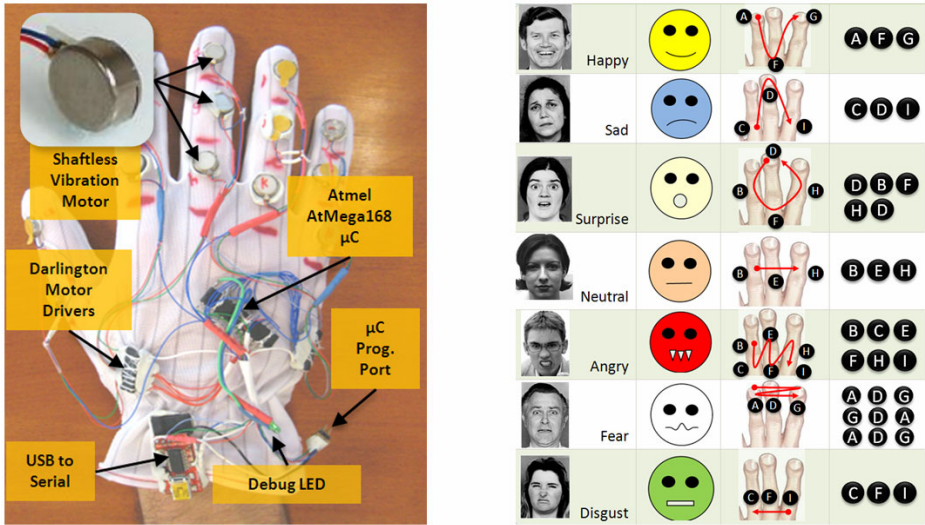


Fig. 4. Vibrotactile glove for conveying facial expressions and the vibrotactile mapping

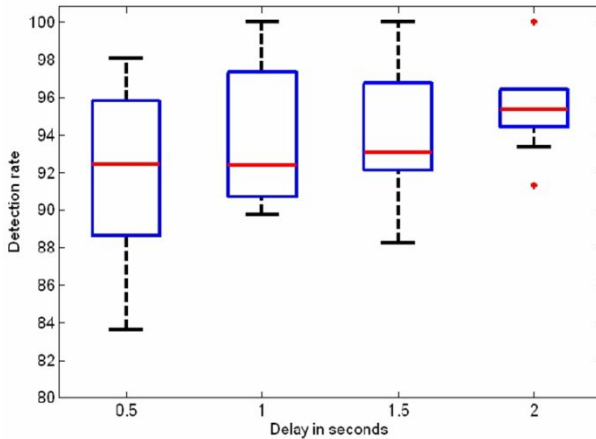


Fig. 5. Box plot analysis of the body rock detection component

to the above experiment, sensors could be placed on the hands, legs, spine etc. to capture and detect stereotypes like head weaving, hand or leg tapping, body swinging etc.

4.2 Conveying Facial Expressions through Haptic Icons

The primary goal of the experiment was to determine how well participants were able to recognize seven haptic patterns, corresponding to seven basic expressions, on the haptic glove. The experiment was conducted with one individual who was blind and 11 other participants who were sighted, but were blindfolded. Fig. 6(a) shows the

confusion matrix of communicating seven haptic expression icons. The diagonals correspond to the recognition accuracies and the off-diagonal elements represent the confusion between expressions. It can be seen that the participants were able to recognize the expressions with an average accuracy of 90%. Vibrotactile pattern corresponding to Sad had the worst performance of 80% while Fear and Disgust had a performance close to 98%.

Fig. 6(b) shows the average time taken by the subjects per expression when they recognized the haptic patterns correctly (cyan), and when they misclassified them (red). The bar graph shows excess or shortage of response time around the mean value. It can be seen that correct identification happened in just over a second (1.4s). When the subjects were not sure of the haptic pattern, they took more time to respond (mean of 2.31s). Fear, which had the highest performance in terms of recognition rate also had the least time of response. The results are promising and work is in progress to extend the glove to convey more complex facial movements allowing users to make the categorization on the expressions using their cognitive faculties.

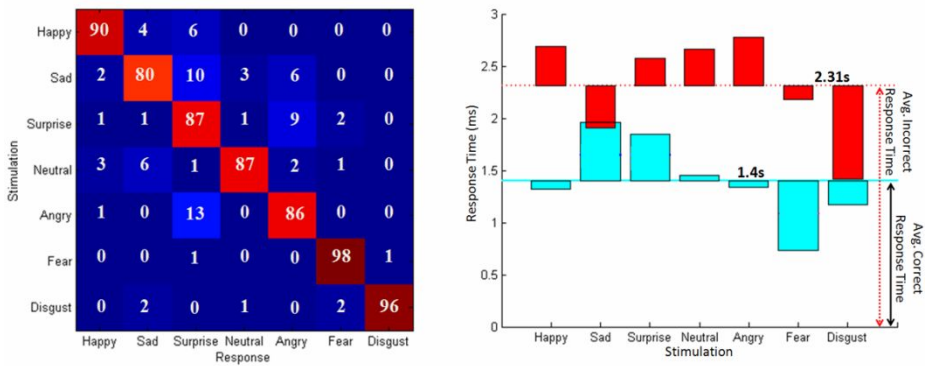


Fig. 6. Confusion Matrix for the seven expressions. (b) Response time for each expression.

5 Conclusion and Future Work

In this paper we have identified an important assistive technology area relating to the social interaction needs of individuals who are visually impaired. We have identified the important issues in social communication and demonstrate how wearable technology could offer enrichment technologies. Specifically, two promising solutions relating to the two primary needs of the target population are discussed. In the future, work will be progressed towards delivering subtle facial and body based non-verbal cues to a user who is visually impaired. More experiments with the target population will be carried out to measure the efficacy of the overall system.

Acknowledgments. We like to thank Stephen McGuire, Troy McDaniel, Shantanu Bala, Jacob Rosenthal, Narayanan Krishnan and Nathan Edwards for their participation in the design, development and testing of the above technologies. We also like to

extend out thanks to Terri Hedgpeth and RaLynn McGuire at the Disability Resource Center, Arizona State University for their guidance for the Social Interaction Assistant project.

References

1. Knapp, M.L., Hall, J.A.: *Nonverbal Communication in Human Interaction*. Wadsworth Publishing, Belmont (2005)
2. Krishna, S., Balasubramanian, V., Black, J., Panchanathan, S.: Person-Specific Characteristic Feature Selection for Face Recognition. In: Boulgoris, N.V. (ed.) *Biometrics: Theory, Methods, and Applications* (IEEE Press Series on Computational Intelligence). Wiley-IEEE Press (2009)
3. Shinohara, K., Tenenber, J.: A blind person's interactions with technology. *Communications of the ACM* 52, 58 (2009)
4. Jindal-Snape, D.: Use of Feedback from Sighted Peers in Promoting Social Interaction Skills. *Journal of Visual Impairment and Blindness* 99, 1–16 (2005)
5. Réhman, S.U., Liu, L.: Vibrotactile Rendering of Human Emotions on the Manifold of Facial Expressions. *Journal of Multimedia* 3 (2008)
6. Krishna, S., Colbry, D., Black, J., Balasubramanian, V., Panchanathan, S.: A Systematic Requirements Analysis and Development of an Assistive Device to Enhance the Social Interaction of People Who are Blind or Visually Impaired. In: *Workshop on Computer Vision Applications for the Visually Impaired, CVAVI 2008* (2008)
7. Krishna, S., McDaniel, T., Panchanathan, S.: *Embodied Social Interaction Assistant*, TR-10-001, Tempe, USA: Arizona State University (2010), <http://cubic.asu.edu/d6/content/embodied-social-interaction-assistant>
8. McDaniel, T., Krishna, S., Balasubramanian, V., Colbry, D., Panchanathan, S.: Using a haptic belt to convey non-verbal communication cues during social interactions to individuals who are blind. In: *IEEE Intl. Workshop on Haptic Audio visual Environments and Games, HAVE 2008*, pp. 13–18 (2008)
9. Krishna, S., Little, G., Black, J., Panchanathan, S.: A wearable face recognition system for individuals with visual impairments. In: *Proc. 7th Intel ACM SIGACCESS Conf. on Computers and accessibility*, pp. 106–113 (2005)
10. Eichel, V.J.: A taxonomy for mannerism of blind children. *J. of Visual Impairment & Blindness* 73, 167–178 (1979)
11. Krishna, S., Bala, S., McDaniel, T., McGuire, S., Panchanathan, S.: VibroGlove: An Assistive Technology Aid for Conveying Facial Expressions. In: *Proc. 28th Intl. Conf. extended abstracts on Human factors in computing systems* (2010)

Clothes Matching for Blind and Color Blind People

Yingli Tian and Shuai Yuan

Electrical Engineering Department
The City College of New York
New York, 10031
{ytian, syuan00}@ccny.cuny.edu

Abstract. Matching clothes is a challenging task for blind people. In this paper, we propose a new computer vision-based technology of clothes matching to help blind or color blind people by using a pair of images from two different clothes captured by a camera. A mini-laptop or a PDA can be used to perform the texture and color matching process. The proposed method can handle clothes in uniform color without any texture, as well as clothes with multiple colors and complex textures patterns. Furthermore, our method is robust to variations of illumination, clothes rotation, and clothes wrinkles. The proposed method is evaluated on a challenging database of clothes. The matching results are displayed as audio outputs (sound or speech) to the users for “match (for both color and texture)”, “color match, texture not match”, “texture match, color not match”, or “not match (for both color and texture)”.

Keywords: Computer Vision, Clothes Matching, Color Matching, Texture Matching, Blind, Color Blind.

1 Introduction

Based on the 2002 world population, there are more than 161 million visually impaired people in the world today, of which 37 million are blind [1]. In everyday life, people need to find appropriate clothes to wear. It is a challenging problem for blind people to find clothes with suitable color and texture. Most blind people manage this problem through the following ways: 1) Through help from their family members; 2) through using plastic Braille labels or different types of stitching patterns which are tagged on the clothes to represent different colors and appearances [16]; 3) through choosing clothes with simple colors.

In this paper, we develop a computer vision-based method to detect if a pair of images of two clothes matches for both texture and color. The image pair is captured by a wearable camera which is connected to a computer or a PDA. To our knowledge, there is no device on the market with this function. The function of matching clothes will also benefit people who are color blind. Figure 1 demonstrates the concept for clothes matching. By performing the texture and color matching to a pair of images from different clothes, our algorithm can detect: 1) colors of the clothes; 2) the texture of the clothes; 3) whether the colors match; and 4) whether the textures match. The matching results can be communicated to the user auditorily as “match (for both color and texture)”, “color match, texture not match”, “texture match, color not match”, or

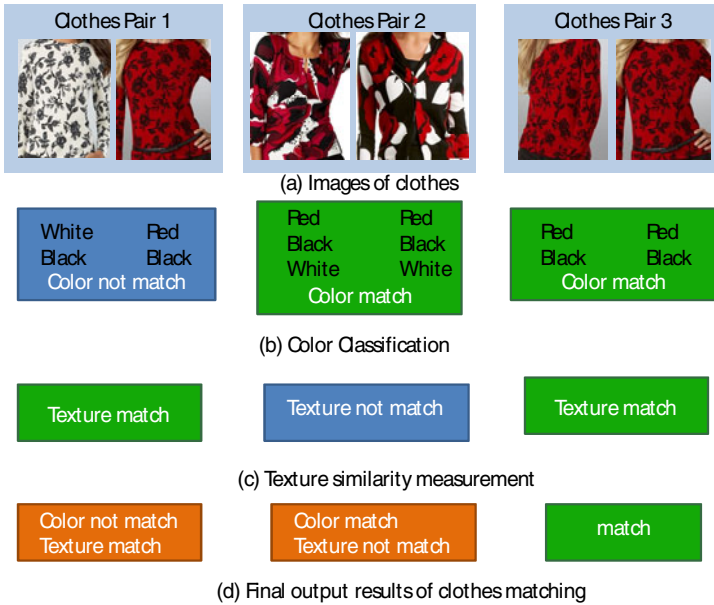


Fig. 1. Matching clothes with multiple colors and complex patterns by using color and texture information. (a) Three pairs of images of clothes. (b) Color classification results. (c) Texture similarity measurement results. (d) Final audio outputs.

“not match (for both color and texture)”. Since most cell phones are with built-in cameras, the algorithm can also be integrated into cell phones.

2 State-of-the-Art

In the current market, there is no device for clothes matching. However, some color identifiers are available but they can only detect primary colors present in a very small region. Figure 2 shows the color identifier manufactured by BRYTECH [2]. This device cannot correctly classify colors of clothes containing multiple colors and complex patterns.

In computer vision and image processing research, many methods were developed for texture and color matching [3-15]. There are three critical issues for successful clothes matching. The first is the issue of color constancy. People perceive an object to be the same color across a wide range of illumination conditions but the actual pixels of an object, which are perceived by a human to be the same color, may have values (when sensed by a camera) that range across the color spectrum depending on the lighting conditions. Secondly, shadows and wrinkles are often part of the texture of clothes and cause errors. Lastly, many clothes have designs with complex patterns and multiple colors. To overcome the above issues, our method is designed to handle clothes with multiple colors and complex patterns by using both color and texture information. In addition, our method can deal with illumination changes, clothes wrinkles, and clothes rotations.



Fig. 2. Color identifier manufactured by BRYTECH [2]

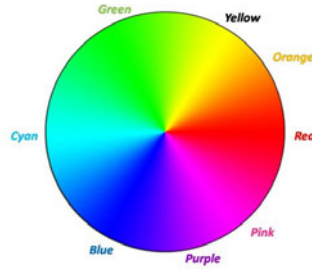


Fig. 3. Basic color space quantization based on Hue for pixels meet constrains of saturation and luminance

3 Methodology for Clothes Matching

3.1 Color Classification and Matching

Our color classifier is based on acquiring a normalized color histogram for each image of the clothes in bi-conic (hue, saturation, luminance) HSI space. The key idea is to intelligently quantize color space based on using the relationships between hue, saturation and luminance. As color information is limited by both lack of saturation and intensity, it is necessary to separate chromatic from achromatic space along surfaces defined by a function of saturation and intensity in the bi-conic space. In particular, for each image of the clothes, the color classifier creates a histogram of the following colors: *red, orange, yellow, green, cyan, blue, purple, pink, black, grey,* and *white*. These colors are selected based on the empirical distribution and our ability to discern. Each image of an article of clothing is first converted from RGB to HSI color space. Next, HSI space is quantized into a small number of colors. If the clothes contain multiple colors, the dominant colors will be outputted.

In our color classification, we first detect colors of “white”, “black”, and “gray” based on saturation S and luminance I . If the luminance I of a pixel is larger than 0.75, and saturation $S < 0.25 + 3(0.25 - (I - I))$, then the color of the pixel is defined as “white”. Similarly, the color of a pixel is “black”, if the luminance I of a pixel is less than 0.25 and saturation $S < 0.25 + 3*(0.25 - I)$. For color “gray”, the saturation S and luminance I should meet following conditions: $0.25 \leq I \leq 0.75$ and $S < I/5 * I$. For other colors (e.g. *red, orange, yellow, green, cyan, blue, purple, and pink*), hue information is employed. As shown in Figure 3, hue is displayed as a 360° color wheel. We define the color “red” between 345°--360° and 0°--9°, “orange” in the range of 10°--37°, “yellow” between 38°--75°, “green” between 76°--160°, “cyan” between 161°--200°, “blue” in the range of 201°--280°, “purple” between 281°--315°, and “pink” between 316°--344°.

The dominant colors will be communicated auditorily to the blind user. As shown in Figure 4, both Figure 4(a) (with texture) and 4(b) (without texture) show examples of clothes in color of light blue. However, the user will not be able to figure out whether the white and blue colors are mixed together only based on the colors classification results. To avoid this kind of confusions, we detect if the clothes have texture patterns by using the method described in Section 3.2.

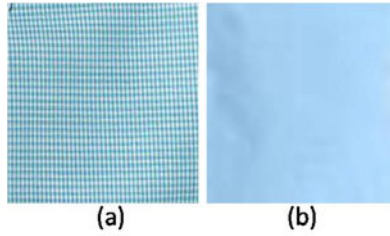


Fig. 4. (a) An example of clothes with texture and mixture colors (64.7% cyan, 27.6% white, 4.6% gray, 3% blue); (b) an example of clothes without texture but with mixture colors (9% blue and 90% white)

3.2 Texture Detection

Figure 5 shows the basic method to detect if the clothes have texture patterns. Based on the color classification results, if there is only one dominate color, the input image of clothes has no texture. Only for the images with multiple dominate colors, we continue to check if the multiple colors are caused by texture patterns. To detect if an image has texture or not, we first transfer it to gray scale image, and then perform Gaussian smoothing to reduce noises. Next Canny edge detection is applied to detect the edges in the image followed by a morphology processing to remove the edge pixels with small areas. At each edge pixel, we check the neighbor pixels around it by using a 3x3 mask to get the directions of the edge. Along the edge direction, we calculate the color values on the both sides of the neighbor's position in the original image. If the colors from both sides do not match each other, the edge is more likely caused by texture. Otherwise, the edge pixel is removed from further processing. Finally, we calculated the total edge ratio with the image having texture if the edge ratio is larger than a threshold. If both images from the image pair are without textures, we treat them as “texture match”. For images with texture, we will continue to perform texture matching as described in Section 3.3.

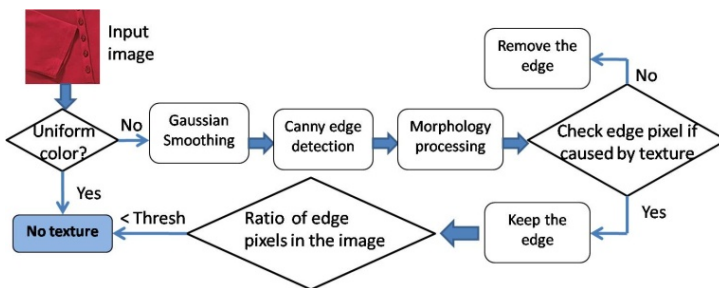


Fig. 5. Flowchart of proposed texture detection method

3.3 Texture Matching

For clothes with complex texture patterns, we develop a new approach for texture analysis by using Radon transform, wavelet features and co-occurrence matrix to

handle illumination changes and rotations. Wavelet features provide spatial and frequency information which can be easily used to analyze texture features. However, wavelet features are stable for lighting changes but sensitive to the orientation of textures. In order to develop methods that work successfully for texture orientation-invariant, we employ Radon transform for estimating the orientation of texture patterns then rotate the image with main orientation of texture as 0° . Histogram equalization is performed to decrease illumination changes. Next, Haar wavelet transform is applied to obtain features on three directions (horizontal, vertical and diagonal). For each wavelet sub images, co-occurrence matrix for gray texture analysis is calculated. Finally, the texture matching is performed based on statistical classification included six features, e.g. mean, variance, smoothness, energy, homogeneity, and entropy.

4 Experiment Results

4.1 Databases

To validate the effectiveness and robustness of our method, we have performed evaluation on 3 datasets: 1) Brodatz album texture database [13]; 2) color and texture matching (CTM) dataset for clothes; 3) texture detection (TD) dataset for clothes. We collect datasets of clothes images with variety of colors and texture patterns. The Brodatz album texture database [13] is a public available standard dataset for texture analysis. We have used it to compare our method to state-of-the-art methods. The CTM database contains 128 images of clothes with complex patterns, multiple colors, and lighting changes. The TD dataset contains 45 clothes images with different patterns and colors. Among these images, 23 of them have textures and 22 images without texture.

4.2 Color and Texture Matching Results

Texture classification results on Brodatz album dataset: To make the results comparable, we apply exactly the same experimental setting on Brodatz album dataset as in [13]. Since we do not have the source code of [13], we implement their algorithm for the comparison. For Brodatz album dataset, there are 60 different textures and each texture is treated as a class. Our method achieves average classification rate of 97%. The method of Khouzani and Zadeh [13] achieves average accuracy of 96.7%.

Color and texture matching results on CTM dataset: Our CTM dataset contains 128 clothed images with complex patterns and colors. Among these images, 76 of them are matching and 52 images are non matching. In our test, we selected 50 matching pairs and 83 non-matching pairs. For texture matching, the method of Khouzani and Zadeh [13] achieves 67.7% matching rate and our method achieves 82.7% accuracy rate. Table 1 shows that our method achieves the state-of-the-art results on the standard Brodatz album dataset and significantly outperforms existing methods on our CTM dataset containing images of clothes with complex patterns and lighting changes. Table 2 shows the confusion matrix of our method for texture matching on CTM dataset. For color detection and matching, our algorithm achieves correct classification and matching rate at 99.2%. Figure 6 displays some examples of

the results for clothes matching by using both color and texture information. The first row in Figure 6 shows the original image pairs. The second row displays the Radon transform image for texture orientation estimation. The third row and the fourth row demonstrate the wavelet features and the co-occurrence matrix images respectively.

Table 1. Comparison with the state-of-the-art results of texture matching on Brodatz album dataset and CTM dataset

Methods	Khouzani and Zadeh [13]		Our method	
	Brodatz	CTM	Brodatz	CTM
Accuracy	96.7%	67.7%	97%	82.7%

Table 2. Confusion matrix of clothes texture matching by our method on CTM dataset

	Number of matching pairs	Number of non matching pairs
Number of matching pairs	42	7
Number of non matching pairs	16	68

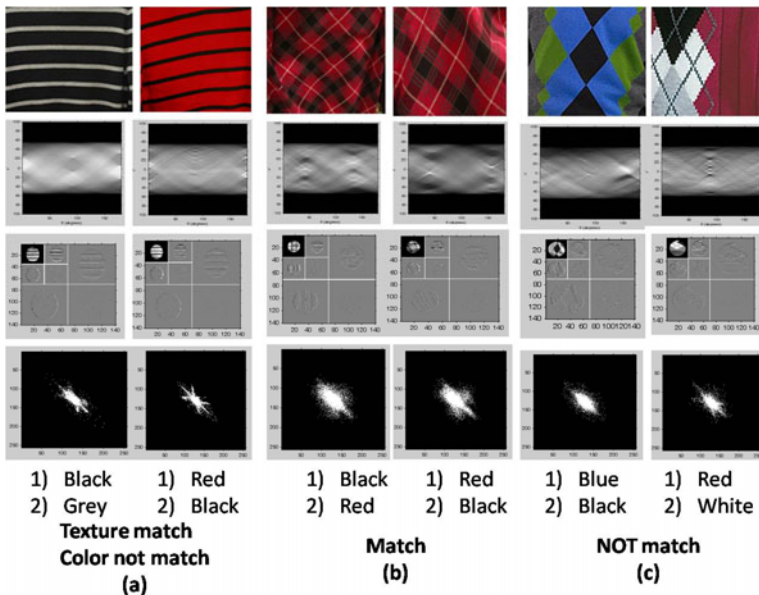


Fig. 6. Examples of results for clothes matching. (a) The clothes images are texture match, but color doesn't match; (b) the clothes images are match for both texture and color; (c) the clothes images are NOT match for both texture and color. The first row shows the original image pairs. The second row displays the Radon transform image for texture orientation normalization. The third row and the fourth row demonstrate the wavelet features and the co-occurrence matrix images respectively.



Fig. 7. Example results of texture detection. The top row shows the clothes without texture and the bottom row shows the example clothes with texture patterns.

Texture detection results on TD dataset: Our TD dataset contains 45 clothes images with different patterns and colors. Among these images, 23 of them have textures and 22 images without texture. In our test the accuracy of our texture detection is 100%. Figure 7 demonstrates some example results of texture detection. Our method can handle wrinkles and lighting changes.

5 Conclusion and Future Work

We have developed a new method to match clothes with multiple colors and complex patterns to assist visually impaired and blind people by distinguishing both texture and color information. To handle complex texture patterns and lighting changes, we combine Radon transform, wavelet features, and co-occurrence matrix for texture matching. Our algorithm for color matching is based on normalized color in HSI color space. We develop a color classifier to detect multiple colors including *red*, *orange*, *yellow*, *green*, *cyan*, *blue*, *purple*, *pink*, *black*, *grey*, and *white*. We also develop a simple edge-based texture detection method. The texture matching is only performed for the images with texture patterns. The proposed clothes matching algorithm is evaluated by three databases. Two of the databases contain clothes images with a variety of texture patterns, colors, and illumination changes. The results demonstrate that our method is robust and accurate for clothes with complex patterns and multiple colors.

The proposed color matching and detection function will also benefit color blind people. This research has the following impacts: (1) It enriches the study of texture matching, and leads to significant improvements over existing methods in handling images with lighting changes and complex texture patterns with different directions; (2) The method developed in this paper provides new functions to improve the quality of life for blind and visually impaired people; and (3) The research will benefit many other important areas including object detection and industry inspection, etc.

Our future work will focus on classifying more colors and transferring the function to cell phones. We will also address the human interface issues for image capture and auditory display of the clothes matching on computers and cell phones.

Acknowledgments. This work was supported by NSF grant IIS-0957016.

References

1. Resnikoff, S., Pascolini, D., Etya'ale, D., Kocur, I., Pararajasegaram, R., Pokharel, G., et al.: Global data on visual impairment in the year 2002. *Bulletin of the World Health Organization* 82, 844–851 (2004)
2. Color identifier manufactured by BRYTECH, <http://www.brytech.com/>
3. Akhloufi, M.A., Ben Larbi, W., Maldague, X.: Framework for color-texture classification in machine vision inspection of industrial products. *IEEE International Conference on System, Man, and Cybernetic* (2007)
4. Caelli, T., Reye, D.: On the classification of image regions by color, texture and shape. *Pattern Recognition* (1996)
5. Charalampidis, D., Kasparis, T.: Wavelet-Based Rotational Invariant Roughness Features for Texture Classification and Segmentation. *IEEE Trans. on Image Processing* 11(8) (2002)
6. Dennis, D., Higgins, W.E., Wakeley, J.: Texture segmentation using 2D Gabor elementary functions. *IEEE Trans. On Pattern Analysis and Machine Intelligence* 16(2) (1994)
7. Do, M.N., Vetterli, M.: Texture similarity measurement using Kullback-Leibler distance on wavelet subbands. In: *Proc. IEEE Int. Conf. on Image Proc., IICIP* (2000)
8. Do, M.N., Vetterli, M.: Rotation invariant characterization and retrieval using steerable wavelet-domain Hidden Markov Models. *IEEE Trans. Multimedia* 4(4) (2002)
9. Van Gool, L.J., Dewaele, P., Oosterlinck, A.: Texture analysis. *Compute Vision, Graphics, and Image Processing* 29 (3) (1985)
10. Haley, G.M., Manjunath, B.S.: Rotation-invariant texture classification using a complete space-frequency model. *IEEE Trans. Image Process.* 8(2) (1999)
11. Hanbury, A., Kandaswamy, U., Adjero, D.A.: Illumination-invariant Morphological Texture Classification. In: *International Symposium on Mathematical Morphology* (2005)
12. Jalil, A., Manzar, A., Cheema, T.A., Qureshi, I.M.: New Rotation- Invariant Texture Analysis Technique Using Radon Transform and Hidden Markov Models. *IEEE, Trans. INF&SYST.* E91-D(12) (2008)
13. Khouzani, K.J., Zaden, H.S.: Radon Transform Orientation Estimation for Rotation Invariant Texture Analysis. *IEEE. Trans. on pattern analysis and machine intelligence* 27(6) (June 2005)
14. Varma, M., Zisserman, A.: Texture Classification: Are Filter Banks Necessary. In: *IEEE Conference on Computer Vision and Pattern Recognition* (2003)
15. Wu, W.R., Wei, S.-C.: Rotation and gray-scale transform invariant texture classification using spiral resembling, subband decomposition, and Hidden Markov Models. *IEEE Trans. Image Process* 5(10) (1996)
16. How Blind People Match Clothing? http://www.associatedcontent.com/article/1788762/how_blind_people_match_clothing.html

A Basic Inspection of Wall-Climbing Support System for the Visually Challenged

Makoto Kobayashi

Tsukuba University of Technology, Japan
koba@cs.k.tsukuba-tech.ac.jp

Abstract. Wall climbing became more and more popular sports among the visually challenged in these days. They can play and enjoy it together with sighted people without any additional rules. However, severely visually impaired climbers still have a problem. That is a difficulty to know where the climbing hold after next one is in advance. A visually impaired climber champion pointed out that to know the positions of holds of two or three steps ahead is very important and that information will be useful to make a strategy. To solve that problem, a basic inspection of support method is conducted. Web camera, a pair of ultra sonic devices and a bone conduction headphone with Bluetooth technology are tested. The results of these tests and comments by climbers suggested that to make a system which support bouldering using general-purpose equipments is available for the visually challenged climbers.

Keywords: wall climbing, visually challenged, bouldering.

1 Introduction

Recently, wall climbing gradually became a popular sport among the visually challenged in Japan. Figure 1 shows that visually challenged persons enjoy climbing. One of the reasons why the climbing had been such a popular sport is that Mr. Koichiro Kobayashi, who is a first prize winner of the world climbing championship for the visually challenged, made an NPO to spread the climbing activity. The NPO named “Monkey Magic” had held many events aim to lecture to beginners since 2005.

Basically, climbing has suitable characters for the visually challenged with reasons as follows: First, they can enjoy it with sighted people, without any additional rules like wearing eye mask or using ringing items. Second, the wall doesn’t move, doesn’t change, and doesn’t attack the climbers. Wall climbing is safety sports for the visually challenged compare with other sports for the blind like soccer or marshal arts, in spite of it seems to be a dangerous one. Third, it is not a much-up type sport. It means that they don’t need to hesitate to stop and to take a rest at any time. In other words, they can keep their own pace. It is important factor especially for the visually challenged players. It can be said same thing to the difficulty level. In usual case, the difficulty level of the climbing is defined by a series of holding stones (they are called as “holds” or “climbing holds”) which is allowed to use. Because of the way to set difficulties, the level can be easily controlled and set as to fit for personal physical ability. Finally, most remarkable point is the fact that even visually challenged person can

support sighted climber as a person to belay. This character can not be seen other sports for the visually challenged.

In spite of having such good features, the wall climbing still has problem for severely visually impaired player, as Koichiro pointed out. That is, difficulty of making a strategy by him/herself. To make a strategy, two or three steps further information of climbing holds is necessary. However, only information they can acquire is the next climbing hold by touching it. Even if climbing holds are connected by a string to make it easy to find next holding stone in the lecture events held by the NPO mentioned above, the player can not get information of holding stone after next one.

To solve that problem, we attempted to develop a support system which gives further information of climbing holds to the climber.



Fig. 1. Climber and belaying two persons are all partially sighted. Visually impaired person can enjoy climbing a wall and they also can support the climber. Strings put on the wall indicate climbing holds which should be grabbed. The visually impaired climber can find a next hold by touching and tracing these strings.

2 Inspection of How to Support

2.1 Interview with Climbers

Before designing the system, interviewing with Mr. Koichiro and his friends was conducted. The comments by the interview showed that they do not want to carry on heavy gadgets like small computer or camera or scanning device when they climb a wall and they do not want to use such kind of support system if it will be expensive one. Though there is another way to support climbing by visually challenged using expensive special device like a Brainport[1], they feels that carrying such device diminishes sportive feelings. Of course it is valuable research to develop special device, however, we made our approach have more practical way.

They also commented that it is not good solution to make special infrastructure like plugging electrode or some ID tags to climbing holds or to the wall. If special wall for the visually challenged will be prepared, they can not enjoy other climbing wall except it.

By these comments, we recognized that the support device should be light-weight if the climber have to carry it and should be inexpensive one.

2.2 System Design

After that, proposed support system was designed considering the results of the interview. The system was designed as to be based on a normal computer to curb the cost. It comprised of a device for detecting position of the climber and a device for providing information. The computer stores position data of climbing holds to be grabbed in advance and the climber do not carry it. It calculates a relative directions and relative intervals between the climber and each holds, using climber position which is detected by certain device. Then, the computer transmits the information to the climber through the providing device.

As for the provide information and device, we chose verbal sound information and bone conduction headphone as a first step. Verbal sound information is already used as supporting visually challenged now and the bone conduction headphone is known as to be suitable for the support system for the visually impaired person because it does not close the ear[2].

2.3 Testing a Web Camera

On the other hands, there are several options on selecting the device for detection. One of these options is a camera. Setting it on the ground and capturing images of a climber and adopting appropriate image processing to detect his/her position seems to be good solution in this case. Because the climber does not needs to carry anything. In addition to it, so-called “Web Camera” has very reasonable price in these days.

To check an ability of a Web camera, we set a camera (logitech QCAM-200R Orbit AF) in front of the climbing wall with three meters distance and captured video image. Maximum resolution of this camera is 1600 pixels by 1200 pixels and this also equips pan and tilt functions. After surveying several climbing video, we thought as follows. Up to about three or four meters high, the resolution is enough and it is

considered to be able to detect the position of the climber with appropriate image processing and able to calculate the distance to climbing holds. However, when the climber move upper than five meters high, the image of the climber is too small to detect. In that case, tilting the camera up is required and it means that the tilting angle information is needed for calculation. It also increase relative angle between projection plane and the wall plane and that causes producing error for calculating. Of course using more expensive camera with optical zoom will solve these problems, but it is not fit for our approach to construct practical system.

2.4 Testing a Pair of Ultrasonic Devices

After testing the Web camera, we tried to take another way to detect a position of the climber. Assuming the climber found climbing holds correctly, it is able to give next and one after next position of holds to the climber only by height information of him/herself. From that view point, device which detects height of the climber was made and tested. Considering safety and fabrication cost, ultrasonic device is chosen as the way to measure the distance.

The left photograph of Figure 2 shows a transducer unit which emits ultrasonic signals. The device is hung on a harness and its weight is less than 94 grams including double AAA batteries, which is considered as light enough. The right photograph shows a receiver unit and a PC on the ground. The receiver and the PC are connected using radio wave. (Therefore, the PC does not need to put on the ground actually.)



Fig. 2. A pair of ultrasonic device which measures distance between each unit. Left photo shows a transducer unit hung on a harness a climber wears. Right photo shows a PC and a receiver unit on the ground. The receiver unit and the PC connected using radio wave.

Using these ultrasonic devices, testing session was conducted. A climber wears the transducer unit and climbed a wall which was about twelve meters high. During the climbing, the PC on the ground displays correct distance and climber commented he did not care the weight. The possibility to make support system using ultrasonic device was shown.

However, ultrasonic device is not considered as general-purpose equipment. Actually it is not so expensive one, although ordinary visually challenged can not get it easily.

2.5 Testing Bluetooth Headphone

Bone conduction headphone using Bluetooth technology is tested. The number of PC equipped with Bluetooth interface is increases recently, therefore it is not need to buy special device for establish wireless connection.

In the test session, the tester wore the headphone and climbed wall with over twelve meters high. During the climbing, the sound source is transmitted from a computer placed on the ground. After all, the sound radio reaches until about ten meters high. However, it is not stable after that height. Because originally Bluetooth is short range wireless standard and the specification of Bluetooth Class 2 defines maximum distance to reach is about ten meters. Using Bluetooth Class 1 devices is more powerful and should be stable, but there is no headphone satisfies the specification.

After these basic inspections, it was clear that up to ten meters high, the combination of ultra sonic device and Bluetooth wireless headphone has possibility to make support device for wall climbing, though ultrasonic device is treated as special hardware.

3 Applying Support System to the Bouldering

During the basic inspections, we could discuss with climbers about these devices and our approaches. In response of the result of testing each devices and comments by

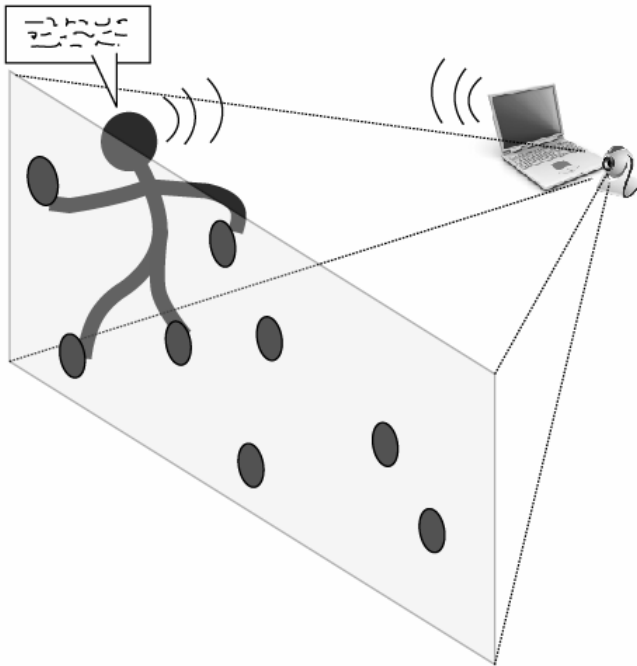


Fig. 3. An overview of proposed support system for bouldering. Web camera captured climber image including whole target wall. Image processing program tracks template marker stuck on the back of the climber and verbal information of next hold and after next hold is given through a Bluetooth headphones or speakers.

climbers, our viewpoint was slightly changed. At first, we only aimed to apply the system to the lead climbing which is most popular style of climbing. However, it was gradually cleared that it is better to apply to a bouldering to realize practical support system.

The bouldering is a climbing style undertaken without a rope, and the height of the wall is limited. We especially focused on a task with motion of lateral direction. Such Limitation of the applying field makes the system realizable. In this case, Web camera can capture the whole image without tilt, pan and zoom. The radio wave of Bluetooth Class 2 transmits from the computer on the ground to the headphone worn on the climber. Everything seems to work well.

Figure 3 shows the overview of the proposal system. Detection device is normal Web camera and image processing program utilizes OpenCV library[3]. Before using this system, appropriate position data to give holds information have to be prepared. After that, operator set template image of target marker which is stuck on the back of the climber. The template image region is displayed by square line which is moved by pressing arrow keys. The width of the square region is controlled by pressing “W” key or “w” key and the height of the region is controlled by pressing “H” key or “h” key. After setting template image region using those key operations, template matching program based on OpenCV library starts. The program tracks the position of the template image and when the position comes certain point set in advance, information of next hold and after next hold is talked by speech synthesizer. That information contains relative direction and relative distance of these holds.

4 Future Works

To operate proposed supporting system for the bouldering, it is still needs sighted operator to fix the position of a targeted template image. The main purpose of this research is to develop a support device by which visually challenged can enjoy this sport totally only by themselves. To realize the purpose, the software of the system should be improved in the future.

Acknowledgement

This work was supported by a Grant for Scientific Research from the Ministry of Education, Culture, Sports, Science and Technology (No.20700457).

Reference

1. A Tongue for an Eye: Device Challenges Conventional Vision, News Blief of National Eye Institute (2009), <http://www.nei.nih.gov/news/briefs/weihenmayer.asp>
2. Kawai, Y., Kobayashi, M., Minagawa, H., Miyakawa, M., Tomita, F.: A Support System for Visually Impaired Persons using Three-Dimensional Virtual Sound. In: Proc. Int. Conf. on Computers Helping People with Special Needs, ICCHP 2000, Karlsruhe, pp. 327–334 (2000)
3. Bradski, G., Kaehler, A.: Learning OpenCV. O’Reilly Media, Sebastopol (2008)

Makeup Support System for Visually Impaired Persons: Overview of System Functions

Akihiko Hanafusa¹, Shuri Terada², Yuuri Miki², Chiharu Sasagawa²,
Tomozumi Ikeda², and Teruhiko Fuwa²

¹ Shibaura Institute of Technology, Department of Bio-science and Engineering,
307 Fukasaku, Minuma-ku, Saitama 337-8570, Japan

² Polytechnic University
4-1-1 Hashimoto-dai, Midori-ku, Sagami-hara, Kanagawa 252-5196, Japan
hanafusa@shibaura-it.ac.jp

Abstract. A questionnaire survey carried out among 25 visually impaired women indicates that they feel uncertainty to their makeup. It is observed that there is a need to develop a system to teach them how to apply makeup and to check their makeup. We have been developing a prototype of a makeup support system for visually impaired persons. The system provides information on makeup and imparts knowledge on how to apply makeup. Further, for checking the current makeup condition, an image processing system that can recognize the face and its parts and check for excess lipstick and the shape of the eyebrows has been developed. From a series of input images, the best image for checking can be selected on the basis of the sum of squares of residual errors after an affine transformation. Further, a usability assessment was carried out by considering eight visually impaired women, and the result could achieve a high score related to the content of information on makeup.

Keywords: Makeup support system, Visually impaired persons.

1 Introduction

To confirm whether or not their makeup looks good, visually impaired persons have to either rely on their own perception or consult someone else. In Japan, some cosmetic companies provide information on their products to visually impaired persons and provide tips on how to apply makeup. This information is provided through Web sites, Braille leaflets, or seminars. However, from the result of a survey carried out on 25 women (congenitally visually impaired: 11; acquired vision impairment: 14), we have found that this dissemination of information is not sufficient. Twenty-one of these women usually apply makeup, but two-thirds of them are not satisfied with their makeup application skills and almost half of them find it difficult to apply makeup and are uncertain of the result. Moreover, 90% of these women think that there should be a computer-based makeup support system that teaches them how to apply makeup and that checks whether they have applied makeup properly.

Several virtual makeup systems and makeup simulation systems[1] have been developed thus far, but they are not suitable for use by visually impaired persons. We have been developing a prototype of a makeup support system[2,3,4] and have successfully included a fundamental image processing function for checking the applied makeup, a learning function for learning how to apply makeup, and a user interface in the system. This paper provides an overview of these functions of the system.

2 System Configuration

The developed system consists of a personal computer with speakers for voice output and a USB camera to capture the image of a user's face, as shown in Fig.1. Fig.2 shows the user interface form of the developed system. It consists of a command input part, an image display area, a makeup procedure area, and a makeup information display area. The user interface and the displayed information can be read out by a screen reader, and all the functions can be performed using the keyboard. Therefore, the system can easily be used by a visually impaired person. Functions for taking pictures, checking the applied makeup, and displaying the makeup procedure and information can be selected using the command input part. The images captured by the USB camera and the results of image processing, such as the recognized face and face parts and results of a makeup check, are displayed in the image display area. The system is developed using C++ and Visual Studio.Net (Microsoft Corp.), and the information is displayed on a text file or an html file. Therefore, the information can be updated easily without having to rebuild the software.

2.1 Methods for Presenting Makeup Information

Two methods are used by the system for presenting the makeup information. The first method is for guiding a user on how to apply makeup. The general procedures for the application of makeup are displayed on the dialogs as shown in Fig.2 and spoken in a sequence; for example, it could present information on how to "wash the face," "apply lotion, emulsion, and foundation," "shape eyebrows," and "apply eyeliner, eye shadow, and lipstick." The user can choose to either repeat a particular procedure or go to the previous or next procedure. Therefore, users can learn how to apply makeup interactively while actually performing the procedures. The other method provides makeup information in a Web page format. Therefore, a user can acquire the information her accustomed way when she uses a Web Browser. The information presented could either be general information on cosmetics and the procedure to use them or information on makeup for special occasions or places such as ceremonial occasions, weddings, and office; the information could even pertain to makeup used for creating certain looks such as cute, active, young, and adult. Further, the recommended makeup application methods for visually impaired persons and their experiences can be displayed. Total number of created pages is 45 and they are written in the XHTML format following the rule of Web accessibility. Precise explanations

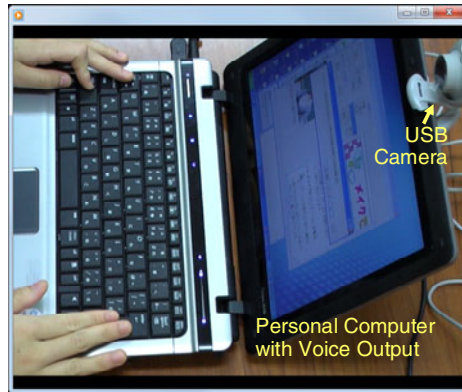


Fig. 1. Configuration of makeup support system

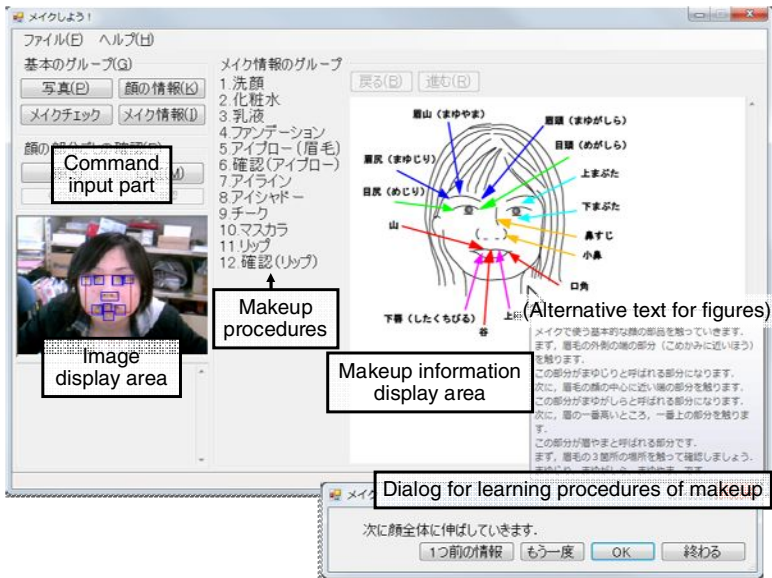


Fig. 2. User interface form of the developed system

on all the figures on the pages are either provided in the body of the page or separately as alternative text. All pages have been checked to ensure that there is no possibility of their being misread by the screen reader and that there is no violation of Web accessibility; for this, the checking tool WebInspector provided by Fujitsu Corp was used.

The system was evaluated, as shown in Fig.1, by considering eight visually impaired women in the age group of 20–70 years; a performance test and a questionnaire were used for the evaluation. Informed consents were obtained from

all the subjects. In the performance test, the time taken for obtaining the desired information and the number of key inputs required for this purpose were recorded. The questionnaire consisted of 21 questions, including seven evaluation factors — operability, understandability of the configuration, listenability, response, favorability, reliability, and usefulness; the questions were marked on a scale of 1 (lowest) to 5 (highest). Compared to the responses of the developer (reference subject) with no visual imparity who operated the system with eyes closed and with the help of voice navigation, the actual subjects of the study required on average 2.9 times the time taken by the reference subject and 3.6 times the number of key inputs provided by the reference subject, to complete the tasks. The average scores of the questionnaire were 2.9 (operability), 3.8 (understandability of the configuration), 4.5 (listenability), 4.3 (response), 4.3 (favorability), 4.3 (reliability), and 4.0 (usefulness). Further, some of the women felt that “information on special occasions or based on many viewpoints is useful.” In particular, the subjects in the age group of 20–30 years gave more than 4 points even for operability, and these subjects required only 1.5 times the time taken by the reference subject. These results show that the developed system provided useful content. However, the operation of the system needs to be improved further, especially for elderly persons who are not skilled at operating a computer.

2.2 Image Processing Method for Makeup Check

The system helps in checking the current makeup condition. The basic method for the makeup check involves a comparison of the current image of the face with a precaptured image of the face before the makeup application or with a satisfactory makeup application. We developed the image processing algorithm using MATLAB (MathWorks Inc.) and incorporated it in the above-mentioned system developed using Visual Studio.Net (Microsoft Corp.). In order to check the applied makeup, it is necessary to recognize the face and its various parts[5,6]. Recently, the concept of face recognition has become very popular and has been used in digital cameras or mobile devices[7,8]. Our system uses a hierarchical face recognition method involving the following three stages:

- 1) The flesh-colored area is extracted from the captured image.
- 2) The flesh-colored area is searched using a coarse template of the face. The current face image resolution is reduced to that of the template for the search process.
- 3) Areas having the required feature points are roughly defined using the face area information. The feature points of the nostrils and the edges of the eyes and the mouth, as shown in Fig.3, are searched by using fine templates in the assumed area. Considering that users are visually impaired, the left and the right edges of the eyes are used as the feature points, instead of the iris.

Template matching[9,10] is performed with a normalized correlation method because this method is rather robust against intensity variations. The average face is used as a template, as shown in Fig.3; this template has been designed on the

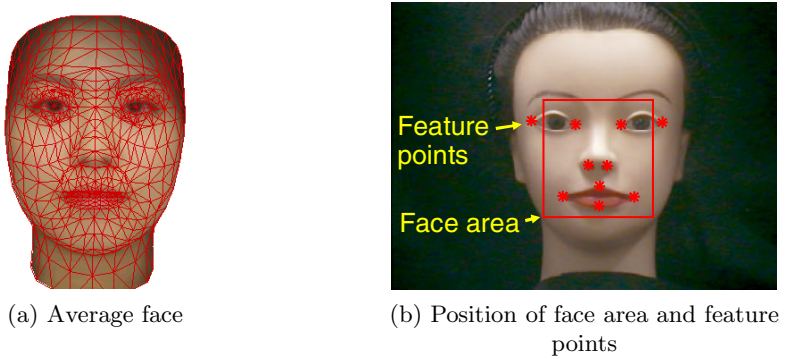


Fig. 3. Average face and position of feature points in the pre-captured image

basis of photographs of the faces of 21 women (age: 18–53 years) who permitted us to use their photographs. The open source software “Face Tool” developed by the Harashima–Naemura Lab in Tokyo University is used for determining the average face.

When comparing the positions of the feature points in the current and pre-captured images, the former is deformed to match the latter using an affine transformation. However, it is impossible to deform the image when the face is turned left, right, up, or down relative to the pre-captured image. Therefore, an affine transformation equation (equation (1)) that is limited to translation, rotation, and magnification is used. The coefficients of the equation are calculated using the least square method and the coordinates of the feature points, and the goodness of image for the transformation can be evaluated by the residual sum of squares (equation (2)), as shown in Fig.4. Hence, the best image for comparison can be selected from a series of input images. Finally, the difference between the abovementioned two images is compared to check the applied makeup.

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} a & b \\ -b & a \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} c \\ d \end{bmatrix} \quad \begin{cases} a = k \cos \theta \\ b = k \sin \theta \end{cases} \quad \begin{cases} c = kt_x \\ d = kt_y \end{cases} \quad (1)$$

$$S^2 = \sum ((ax + by + c - x')^2 + (-bx + ay + d - y')^2). \quad (2)$$

In these equations, k is the magnification coefficient; θ , the rotation angle; (t_x, t_y) , the translation distance; and S^2 , the residual sum of squares.

Thus far, using MATLAB, we have developed functions to check excess lipstick in addition to lip and eyebrow shapes. The shape check system has been implemented on the user interface form shown in Fig.2; however, the implementation of the functions to check excess lipstick and to provide a voice output for the result of the check are remained for the future study.

For checking excess lipstick, an area around the lip is estimated on the basis of the extracted face feature points. As shown in Fig.5, the current image (a) is transformed to image (b) by the affine transformation equation in equation (1) and compared with the pre-captured image (c). A color image of the area is

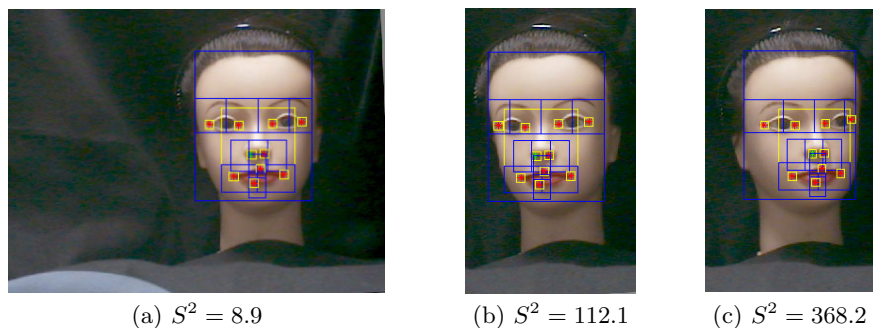


Fig. 4. Examples of S^2 for different face directions

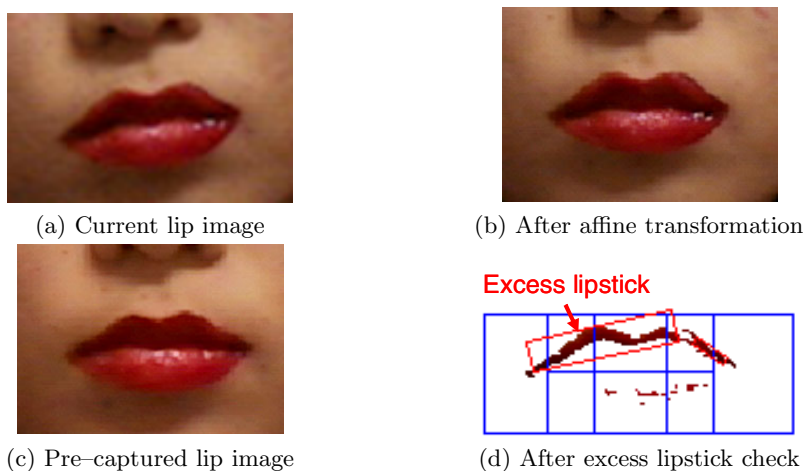


Fig. 5. Example of excess lipstick check

examined in the hue, saturation, and intensity color mode; when the difference in hue color between pixels in the area exceeds the threshold value, the pixels are selected. When the area of a group of selected pixels exceeds the threshold, the area is reported as an area with excess lipstick, as shown in the figure (d). However, the threshold for the difference in hue color would vary depending on the color of the lipstick.

For extracting the contours of an eyebrow, after an average filter and a differential filter are applied in the vertical direction, pixels that have an extreme value in each column are selected. Next, spline approximation is performed on the extracted contour. Four feature points — head, top, tail, and the midpoint of the head and the top — of the approximated spline, as shown in Fig.6, are determined from the spline function and used as an index of the shape of the eyebrow. The example shows a comparison between smooth and sharp eyebrows. Because the image is aligned with the pre-captured image, coordinates can be

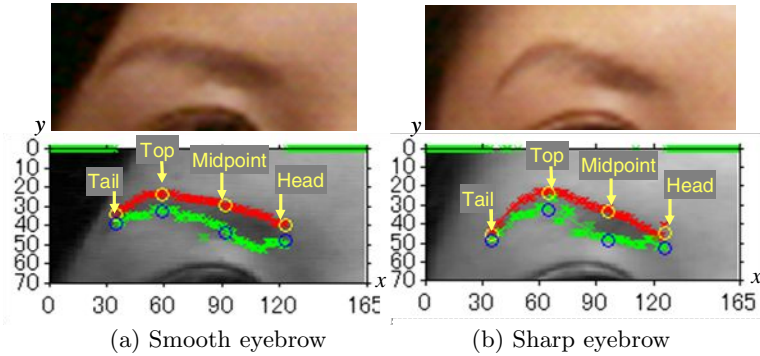


Fig. 6. Example of eyebrow shape comparison

compared directly. The coordinates of the top of the smooth and sharp eyebrows are (59, 24) and (65, 24), respectively, in this case.

3 Conclusion

A prototype of a makeup support system for visually impaired persons was developed. The system provided information on makeup and a function for learning how to apply makeup. Further, for checking the current makeup condition, an image processing system that could recognize the face and its parts and that could check for excess lipstick and the shape of the eyebrows was developed. The result of a usability test carried out on visually impaired subjects showed that the developed system achieved a high score on factors related to the content of information on makeup. However, the operation of the proposed system needs to be improved further, especially for elderly persons who are not skilled at operating a computer. In a future study, more makeup checks will be built into the system and a voice-enabled feedback system will be developed to gauge their usefulness.

Acknowledgments

The authors are grateful to all persons who cooperated to answer questionnaires and evaluate the developed system.

References

1. Furukawa, T., Sakaguchi, Y., Tsukada, A.: Real Time Makeup Simulator. In: Proc. 19th Sensing Forum, pp. 53–38 (2002) (in Japanese)
2. Sasagawa, C., Hanafusa, A., Fuwa, T.: Makeup Support System for the Visually Impaired Person. In: Proc. 2005 JSPE Spring Meeting, pp. 305–306 (2005) (in Japanese)

3. Miki, Y., Sasagawa, C., Hanafusa, A., Fuwa, T.: Makeup Support System for Visually Impaired Persons — Automatic selection of face image that is appropriate for makeup check. In: Proc. 4th Joint Conf. of Life Support Societies, p. 56 (2006) (in Japanese)
4. Terada, S., Hanafusa, A., Ikeda, T., Fuwa, T.: Makeup Support System for Visually Impaired Persons – Usability Assessment. In: Proc. 7th Joint Conf of Life Support Societies, pp. 102–103 (2009) (in Japanese)
5. Funahashi, T., Yamaguchi, T., Tominaga, M., Koshimizu, H.: Facial Parts Recognition by Hierarchical Tracking from Motion Image and its Application. *IRICE Trans. Inf. Syst.* E87–D(1), 129–135 (2004)
6. Liew, A.W.C., Leung, S.H., Lau, W.H.: Lip Contour Extraction from Color Images Using a Deformable Model. *Pattern Recogn.* 35(12), 2949–2962 (2002)
7. Lao, S., Masato, K.: Vision Based Face Understanding Technologies and Their Applications. In: Li, S.Z., Lai, J.-H., Tan, T., Feng, G.-C., Wang, Y. (eds.) *SINO-BIOMETRICS 2004*. LNCS, vol. 3338, pp. 339–348. Springer, Heidelberg (2004)
8. Ijiri, Y., Sakuragi, M., Lao, S.: Security Management for Mobile Devices by Face Recognition. In: 7th International Conference on Mobile Data Management, p. 49 (2006)
9. Brunelli, R., Poggio, T.: Face Recognition: Features versus Templates. *IEEE Trans. on Pattern Analysis and Machine Intelligence* 15(10), 1042–1052 (1993)
10. Kaneko, S., Satoh, Y., Igarashi, S.: Using Selective Correlation Coefficient for Robust Image Registration. *Pattern Recognition* 36(5), 1165–1173 (2003)

WebTrax: Visualizing Non-visual Web Interactions

Jeffrey P. Bigham and Kyle Murray

Department of Computer Science
University of Rochester
Rochester, NY 14627

`jbigam@cs.rochester.edu`, `kyle.murray@rochester.edu`

Abstract. Web accessibility and usability problems can make evaluation difficult for non-experts who may be unfamiliar with assistive technology. Developers often (i) lack easy access to the diversity of assistive technology employed by users, and (ii) are unaware of the different access patterns and browsing strategies that people familiar with a specific assistive technology tool might use. One way to overcome this problem is to observe a person with a disability using their tools to access content, but this can often be confusing because developers are not familiar with assistive technology and tools are not built supporting this use. In this paper we introduce *WebTrax*, a tool that we have developed to support developers who engage blind web users as part of their accessibility evaluation or education strategy. WebTrax helps visualize the process that screen reader users employ to access content, helping to make problems more obvious and understandable to developers.

Keywords: web accessibility, visualization, screen reader, web trails.

1 Introduction

We introduce *WebTrax*, a tool that visualizes the non-visual web browsing of screen reader users to help sighted developers better understand how blind people access web content with a screen reader.

The diverse issues involved in creating accessible web content can make development a frustrating, subjective, and difficult task. Guidelines and standards serve as a valuable starting point for developers hoping to create accessible web content, but many subjective and practical factors influence realized accessibility. Accessibility and usability are often linked [11], and if a resource is not usable then it is in effect not accessible even if it is technically accessible.

In this paper, we explore a different, complementary approach that enables web developers to gain a better understanding of the usability of their own sites. Screen readers are complex software packages that can be intimidating and difficult to understand for those without adequate experience. Even watching a skilled user “drive” a screen reader can be confusing - the screen reader speaks a wide array of content, some derived from the web page it is reading and



Fig. 1. In this visualization from *useit.com*, a user’s attention has been captured by an eye tracker and visually represented as a heatmap over the original page. WebTrax provides analogous visualizations of non-visual attention and usability.

some meant as helpful information for the user. Something as simple as “link ‘Search’” or “Table 4, 3 rows, 2 columns” can be confusing to someone who has not heard it before. Most existing screen readers either do not provide support for highlighting what is being read or provide only minimal support. To understand usability, it would be useful to have a non-visual equivalent of the heatmaps generated by eye-tracking data (Figure 1). WebTrax visualizes the content that is being read in order to help sighted observers make sense of the screen reader experience.

WebTrax is built into the web-based screen reader WebAnywhere [3]. WebAnywhere is unique among screen readers in its ability to be run anywhere - on any platform, from any browser, without requiring special permissions. This means that WebTrax can also be run from any computer, which is an important advantage for a tool that we hope will support learning about accessibility. Educators or WebAnywhere users can demonstrate the non-visual browsing experience to people wherever they happen to be without extensive setup.

1.1 Summary of Contributions

This work makes the following contributions:

- We motivate visualizing non-visual web interactions to help developers improve their understanding of how screen reader users browse through content.
- We describe three visualizations for non-visual web browsing: basic, phosphorous, and heatmaps.
- We present implementations of these visualizations that are integrated into the WebAnywhere non-visual web browser so they are accessible anywhere.



Fig. 2. The basic visualization. (a) The text currently being read is shown above the content in a high-contrast, magnified view to assist those reading along and those with low-vision. (b) The current content being read is highlighted so that users and developers can more easily read along with WebAnywhere.

2 Related Work

WebTrax is designed as both an interactive tool to help web developers better understand the browsing they are observing screen reader users perform and as an educational tool to help teachers convey subtle issues concerning web accessibility and usability.

Accessibility Evaluators: Many accessibility validators have been created to help developers evaluate and improve their web pages. Popular tools include Bobby [17], FAE [5], and WAVE [18]. Developers point these tools at their web pages, and receive back a list of potential accessibility problems. Although a valuable first step in ensuring that one's web page is accessible, evaluators have two primary shortcomings. First, validators cannot evaluate accessibility issues that cannot be detected automatically; for example, evaluation tools can detect if an image lacks alternative text, but they cannot judge if that alternative text is appropriate and informative [4]. Second, usability problems, such as problems with reading order or lack of heading tags or other markup, are difficult to detect automatically. Such usability problems are likely to become more pervasive as web pages become more complex and begin to behave more like applications than static documents [12].

As a fallback, evaluation tools present warnings for content that may have a problem. Evaluation tools often present so many warnings that novice developers might be overwhelmed or be discouraged from fixing anything. While those who are motivated will investigate all warnings, typical developers may be unwilling to investigate each issue, especially when many (or most) will end up not being problems. As an example, WAVE displays a warning anytime a `tabindex` attribute is used within a web page. This attribute can be used correctly to

enforce a meaningful tab order, but when used incorrectly results in a confusing ordering. WebTrax can help visualize how real users might browse this content and expose whether the content is truly a concern.

Simulation: Another approach to evaluation is to provide views of content that simulate what a disabled user might experience. For instance, ADesigner can visually simulate the usability of a web site as either a blind, low-vision, or color-blind person might experience it [14]. Recent extensions add a visual overlay that illustrates reading order [13]. WAVE has recently introduced features designed to help developers appreciate the problems that disabled users might experience when accessing the web. For instance, one option displays a web page in the linear order exposed by screen readers. Although such tools help developers understand how certain groups might experience the web, they miss the personal nature of accessibility and ignore the experience that many users have developed to overcome common issues.

Screening with Assistive Technology: *Screening* is the use of assistive technology to help identify accessibility issues. Henry describes the advantages and limitations of using screening techniques to evaluate accessibility [6]. The advantage is that it may help someone appreciate the experience of someone with different capabilities. A disadvantage, especially when using a complicated software program like a screen reader, is that an inexperienced user may incorrectly associate their own inability to use the tool with accessibility problems. Although valuable, screening also requires expensive screen readers that developers are unlikely to install.

Engaging Users: Mankoff *et al.* compared the results of multiple sighted developers using screening techniques with evaluation by remote blind users [8]. Developers using screen readers found many more problems than did the blind evaluators, although the problems found by blind users were quite accurate. Commonly, when blind users perform an evaluation they focus on the most problematic accessibility problems and sometimes are not able to fully evaluate a web site because the accessibility problems discovered prevent full access to the site [10]. Takagi *et al.* describes the problems of users not knowing what is not accessible to them as a challenge for IBM's Social Accessibility project [15]. WebTrax seeks to help blind web users or experienced educators demonstrate the non-visual web browsing experience to sighted developers.

TrailBlazer is a programming by demonstration system that lets blind web users record, share and replay web tasks [2]. The recordings that they make are expressed in the CoScripter language [7], which uses pseudo-natural language commands that are easy to understand. User actions are recorded as commands like "Click on the 'Search' button" and "enter 'University of Rochester' into the 'Search' textbox". TrailBlazer scripts are shareable, so a user can record themselves navigating through a web page and then share that with a developer in order to convey to how they browsed the content. Unlike WebTrax, however, TrailBlazer does not record information about all of the content that has been viewed by blind web users - it records actions that have been taken but not the

content that has been read. The visualizations supplied by TrailBlazer may allow developers follow along, but, because only one step is highlighted at a time, it can be difficult to get a global view of what the user has done.

3 Visualizing Web Browsing with WebTrax

To use WebTrax, users first load the WebTrax-enabled WebAnywhere and turn on one of the supported visualizations with an associated shortcut key. The blind user or educator driving the interface then browses web content and visualizations are added automatically within the existing WebAnywhere interface.

3.1 Basic Visualization

The basic visualization highlights the current content that is being read (Figure 2). This feature was initially created to allow WebAnywhere to work better for people with low-vision, but we have found that developers appreciate it because it makes following along easier.

3.2 Phosphor Tracking

The *Phosphor Tracking* visualization keeps elements highlighted even after they have been read in order to remind observers of the trail that users have taken

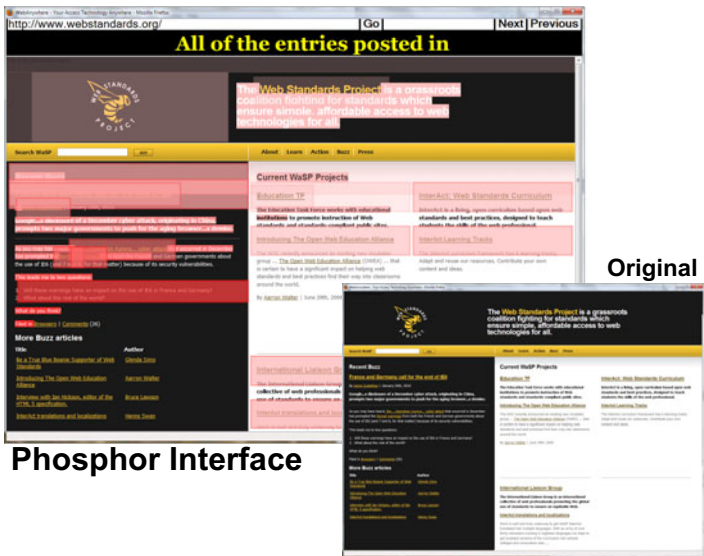


Fig. 3. The phosphor visualization of non-visual web browsing keeps the elements that users visit visible for some time and lets them slowly fade away over time. In this example, we can see that the content on the right actually comes earlier in the reading order than the content on the left because it has faded more as the user has browsed.

through the content of the web page. This visualization was inspired by Phosphor interfaces [1], which are designed to highlight the interface components with which users have interacted and fade slowly away once they are done. These interfaces have been shown to help users remember what they've done and highlight errors.

Our Phosphor Tracking visualization reveals not only the content that has been read but also a sense of the time elapsed since the user visited each part of the page (Figure 3. Content is highlighted as it is read to the user. This highlighting slowly fades over time leaving a trail back to the start of the page.

3.3 Generating Heatmaps

Heatmaps have become a staple in usability as a way to visualize eye-tracking data [9]. Heatmaps display portions of the content where users have looked the most a warmer color (such as red) and portions of the content where users looked the least a cooler color (such as blue). These heatmaps are made possible by expensive and difficult to calibrate eye-tracking technology that records where sighted users are looking.

The third visualization that WebTrax supports is a heatmap over non-visual interactions. This visualization uses two sources of data to generate the heatmap and its color. First, any web page elements that a user has visited is colored.

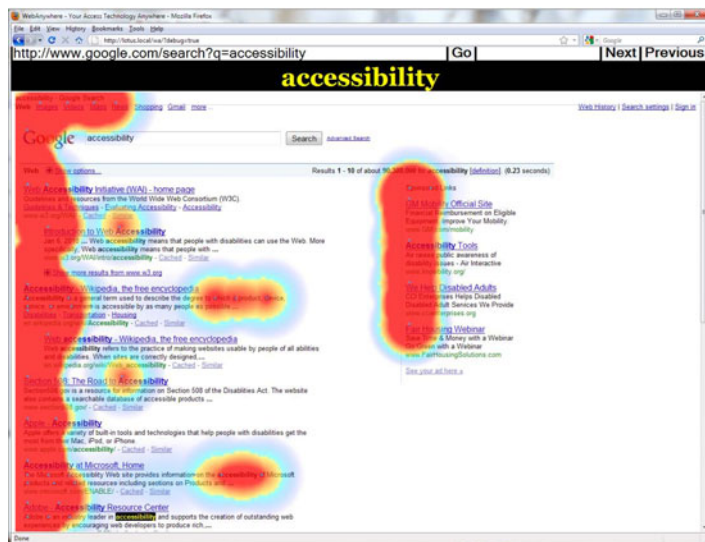


Fig. 4. The heatmap visualization of a user interacting with `google.com` search results shows all of the elements visited by screen reader users with color determined by their displayed interest that content. In this example, we can tell from the heatmap that the user was able to find the main search results, but was distracted at first by the ads on the page. The user was briefly distracted by the options near the beginning of the page, and the summaries of each search result were, for the most part, not fully read.

The color is determined based on how much of the content a user listened to before skipping past the element. For example, if the user listened to an entire heading then that heading would be colored red, if they quickly skipped past it then it would be colored blue, and if they listened to only half of the heading then it would be colored yellow-orange (Figure 4).

The heatmap visualization is created in a transparent Flash movie that has been positioned over the portion of the page that has been read. Each point is assigned a weighted heat value, and its influence is distributed over a fixed radius. Several points can influence the displayed value for a single rendered pixel. The color for a given pixel in the bitmap is determined by the amount of influence that has been exerted over that pixel by all of the elements near it [16]. The values are distributed over a limited hue space in the HSV color representation system. The highest or “hottest” values are red, followed by orange, yellow, and blue for lower or “colder” values.

Heatmaps give observers a global picture of what content is easy to access, and which irrelevant content users must hear before finding what they want. We have found heatmaps are quite adept at identifying the templates of popular web sites - users skip quickly past template information (such as sidebars) causing them to be colored blue, while the main content areas are often colored red.

4 Future Directions

WebTrax is a promising approach for helping web developers better understand the interactions and browsing strategies of screen reader users. We plan to conduct formal users studies to evaluate the efficacy of these visualizations in helping sighted observers better understand how blind web users are browsing through content. Particularly promising applications of this technique include helping to educate web developers and helping web developers evaluate their content.

Because WebTrax is built on top of WebAnywhere, it suffers from the same limitations of the WebAnywhere platform. Specifically, WebAnywhere does not let its users browse local files or files located behind a firewall, which means that developers who want to try WebTrax on their content will need to put the content on a publicly-facing site. These issues are not fundamental limitations of the visualizations presented here, but rather limitations of our current implementation. Many popular screen readers do not provide the programmatic access necessary to support the visualizations introduced here. We hope other screen readers will either expose these visualizations as well or open their APIs to allow third-party tools to do so.

A promising direction for future research is to aggregate the web browsing histories of multiple users together to create more comprehensive visualizations of browsing patterns and problems. The data used to generate these visualizations may also be useful for automatically fixing observed problems by, for instance, making important content that is rarely visited closer to the beginning in the screen reader’s reading order.

5 Conclusion

We have presented and motivated three new visualizations that may help developers and others better understand how screen reader users are interacting with the web pages that they visit. Few tools seek to directly support a collaboration between assistive technology users and developers, but we see this as a promising new approach to teaching developers about the importance and subtleties of accessibility and usability.

References

1. Baudisch, P., Tan, D., Collomb, M., Robbins, D., Hinckley, K., Agrawala, M., Zhao, S., Ramos, G.: Phosphor: Explaining transitions in the user interface using afterglow effects. In: *UIST 2006* (2006)
2. Bigham, J., Lau, T., Nichols, J.: Trailblazer: Enabling blind users to blaze trails through the web. In: *IUI 2009* (2009)
3. Bigham, J.P., Prince, C.M., Ladner, R.E.: Webanywhere: A screen reader on-the-go. In: *W4A 2008* (2008)
4. Craven, T.C.: Some features of alt text associated with images in web pages. *Information Research*, 11
5. Firefox accessibility extension. Illinois Center for Information Technology (2006)
6. Henry, S.: Just Ask: Integrating Accessibility Throughout Design. Lulu.com, London (2007)
7. Leshed, G., Haber, E.M., Matthews, T., Lau, T.: Coscripiter: Automating & sharing how-to knowledge in the enterprise. In: *CHI 2008* (2008)
8. Mankoff, J., Fait, H., Tran, T.: Is your web page accessible?: a comparative study of methods for assessing web page accessibility for the blind. In: *CHI 2005* (2005)
9. Nielsen, J., Pernice, K.: *Eyetracking Web Usability*. New Riders Press, Indianapolis (2009)
10. Petrie, H., Hamilton, F., King, N., Pavan, P.: Remote usability evaluations with disabled people. In: *CHI 2006* (2006)
11. Petrie, H., Kheir, O.: The relationship between accessibility and usability of websites. In: *CHI 2007* (2007)
12. Roadmap for accessible rich internet applications (wai-aria roadmap). World Wide Web Consortium (2007), <http://www.w3.org/TR/WCAG20/>
13. Sato, D., Kobayashi, M., Takagi, H., Asakawa, C.: What's next? a visual editor for correcting reading order. In: Gross, T., Gulliksen, J., Kotzé, P., Oestreicher, L., Palanque, P., Prates, R.O., Winckler, M. (eds.) *INTERACT 2009*. LNCS, vol. 5726, pp. 364–377. Springer, Heidelberg (2009)
14. Takagi, H., Asakawa, C., Fukuda, K., Maeda, J.: Accessibility designer: visualizing usability for the blind. In: *ASSETS 2004* (2004)
15. Takagi, H., Kawanaka, S., Kobayashi, M., Sato, D., Asakawa, C.: Collaborative web accessibility improvement: Challenges and possibilities. In: *ASSETS 2009* (2009)
16. Wand, M.P., Jones, M.C.: Comparison of Smoothing Parameterizations in Bivariate Kernel Density Estimation. *Journal of the American Statistical Association* 88(422) (1993)
17. Watchfire bobby, <http://www.watchfire.com/products/webxm/bobby.aspx>
18. Wave web accessibility evaluation tool (2009), <http://wave.webaim.org/>

A Context-Based Grammar Generation in Mixed Initiative Dialogue System for Visually Impaired

Jaromír Plhák

Faculty of Informatics, Masaryk University
Botanická 68a, 602 00 Brno, Czech Republic
xplhak@fi.muni.cz

Abstract. In this paper, we present some principles of designing mixed-initiative dialogue systems suited to the needs of visually impaired users related with dialogue based web page generation. These principles have been implemented in the online BrowserWebGen system that allows users to generate their web pages. The primary user interface is implemented as a dialogue system with implicit confirmation. The users are able to enter one piece of semantic information at each step of a dialogue interaction. Second part of the application allows the user to activate mixed initiative dialogue interface in each dialogue state. This interface is called Dialogue Prompt Window and it provides an additional support to the users as well as control over the application in natural language.

Keywords: web page creation, dialogue system, grammar, accessibility, visually impaired.

1 Introduction

Dialogue systems that support mixed initiative dialogue communication require a robust dialogue manager. On the other hand, these systems achieve high user satisfaction and support more effective interaction with the application.

In the beginning of the dialogue, the initiative is usually taken by the user answering the open-ended prompt "How may I help you?" (HMIHY). The system adapts its strategy by evaluating the user's inputs within the dialogue manager. Some approaches to dialogue management are based on context representation [1], dialogue tree model with dialogue objects [2] or exploitation of neural networks [3]. Most of the current dialogue systems integrates dialogue managers that support the mixed initiative communication (e.g., the CommandTalk system [4], TOOT [5] or MIMIC [6]).

Designers of spoken dialogue systems develop robust human-machine interfaces using dialogue context and state dependency that affects the improvements of the language models and connection between natural language analyzer and dialogue manager [7]. In most cases, n-gram language models are used for speech recognition [8,9]. These principles allow us to specify complex and efficient grammars for user inputs. In our approach, we exploit these grammars for web page generation.

For many users, creating a web presentation is a nontrivial task. Hence, many online and offline authoring tools like WYSIWYG editors, Content Management Systems (CMSs) or Wizards have been developed to simplify the creation of web pages.

For the visually impaired users, handling of these applications is usually complicated. The WYSIWYG editors requires a pointing device for inserting and editing pictures and other non-textual data. Also in the well-known CMSs, the accessibility difficulties have been addressed (see e.g. [15]). Moreover, they are usually very complex and some training with the user is required. Wizards are more suitable for the blind users, but they offer only a small set of page types and many of them do not assure the accessibility and validity of the resulting web pages.

In [10,11] we have proposed a fully accessible dialogue based system called WebGen. It enables creating web site step-by-step, using forms for data acquisition and web page templates. It also guarantees accessibility of the generated web presentation. Similar principles have been used for developing our application presented in the next section. This system is extended using the Dialogue Prompt Window that is described in Section 3. Some examples of dialogues demonstrating additional functionality are included as well.

2 BrowserWebGen

BrowserWebGen (BWG, [16]) system is an online application that allows the users to create web presentations using web page forms that simulate dialogue. Basic principles are similar to WebGen system, as they were specified in [11,12,13] and differ just in their interface implementation. BWG uses a web browser and screen reader and does not require installation of a special software or knowledge of web technologies, such as HTML or CSS. As opposed to the current CMSs, the user is able to operate the application intuitively without any additional knowledge of the system structure and its modules. The process of web page generation fulfils the Web Content Accessibility Guidelines (WCAG, [14]) as well as XHTML [17] recommendations and is not dependent on graphical interface.

In BWG, the users are allowed to choose their language and identify themselves. Then, they enter the type and name of the presentation. Consequently, the types of information that appear in the presentation are specified. Graphical objects are processed in cooperation with the GATE dialogue system [18,19]. An illustrative example of one step in the generation process is given in Figure 1.

In the final step, the required textual data are specified. BWG system offers several graphical layouts for each type of page. An archive which contains the presentation and which can be uploaded to another server is provided. In Figure 2, we present an example of generated web presentation.

BWG employs the screen reader mode for reading web forms. This mode is well known in the blind computer users community. Therefore, we can safely use controls like checkboxes, radio buttons, menus or text inputs [20]. The user

Contact information

Enter your e-mail

E-mail is john_obrien@gmail.com.

Now, you should enter your address step-by-step. Enter the name of the street:

Name of the street is Central Avenue.

Enter the house number

House number is 1056.

Enter the city:

Your city is Somewhere

Enter the post code. This information is mandatory.

Post code is 10500.

Enter the country:

The country is Canada.

Enter your mobile phone

Your mobile phone is +420174417441.

Enter your ICQ number

ICQ is 174417441.

Your Skype login

Skype login is i_obrien.

Enter your MSN name:

MSN name has not been entered.

Enter your IRC ID please.

Enter your jabber login please.

Enter your AIM login please.

Fig. 1. BrowserWebGen forms for acquiring user inputs

navigates through the dialogue with the standard tabbing navigation and pre-defined access keys. The system prompts the requested pieces of information in linear order and the user is allowed to enter single piece of information at each step of the dialogue interaction. Accessibility of this approach have been verified by testing with the blind students. In order to allow the experienced users to control the application efficiently, we propose the extension of this application based on the Dialogue Prompt Window specified in the next section.

3 Dialogue Prompt Window

Dialogue Prompt Window (DPW) is an additional part of the online applications based on the web page forms. DPW can be activated in any dialogue state and provides the option of operating the application using natural language.

The user is allowed to change the interface without a correction of dialogue initiative. The system translates the label of the active control in the web form

John O'Brien - My first page

About me Curriculum Vitae **Guestbook**

Kontaktní informace:
Jméno: John O'Brien
Email address: john_obrien@gmail.com
Address: Central Avenue 1056, Somewhere 10500, Canada
Cellular phone: +420174417441
ICq: 174-417-441
Skype: j_obrien

Hobbies and interests:

- swimming
- trips
- pop music

Knowledges:

- German - advanced level
- basic computer skills

Birth information:
Date of birth: 1. 1. 1980
Place of birth: Somewhere

Generated using [WebGen system](#). Last updated: 11. 2. 2010

[Czech Blind United](#)
[BrailleNet](#)
[TyfloCentrum Brno](#)

Fig. 2. The web page generated by the BrowserWebGen system

into the system prompts and the appropriate grammar is generated automatically from the type of the control. These grammars are written manually to decrease the number of misunderstandings. Dialogue manager resolves one piece of information at each dialogue step. The user is encouraged to ask for additional help, repetition of the last system prompt, suggestion of possible answers, etc.

3.1 Dialogue Manager

In this section, we describe an extension of the dialogue manager that allows the user to control the application using arbitrary utterances in natural language. The system prompts the HMIHY question and the utterance is analyzed using the grammar induced from the automatically generated grammars or their manually defined equivalents. Nonetheless, complex grammar compositions generate many misunderstandings and collisions, especially when a large number of functions is supported.

Therefore, we propose a principle that specifies appropriate grammars used in the general grammar for the actual dialogue exchange. It is based on the assumption that the user more likely changes piece of information connected to the actual context than the piece of information presented in the different part of

the system. This principle is implemented in the dialogue manager that process the user's input by this complex grammar. The actual context and the complex grammar are reevaluated when the user's request is completely resolved.

Control functions are divided into two groups. The first group is available in every dialogue state. For example, the user is allowed to ask for context-based help anywhere as well as for general help which is divided into topics. Context-dependent functions are in the second group. They allow the user to add, edit or remove the specific piece of information feasible in the actual context.

The dialogue manager proves the correctness and completeness of the entered prompt. Mandatory piece of information are filled in using the confirmation and correction sub-dialogues. As presented in Example 2, basic inference rules are integrated into the dialogue manager as well.

3.2 Example of Functionality of BWG Dialogue Manager

Three mixed initiative dialogues are used in the following examples.

(After choosing language and their identification the user initiates DPW.)

S: How may I help you?

U: I would like to create personal page called My first page.

S: OK.

(The system changes the type of presentation to "personal page" and name of presentation to "My first page". The state of the dialogue is reevaluated and the user is returned to the application because they did not specify that dialogue have to be performed using a natural language.)

Ex. 1. Example of a dialogue between a user and BWG system (processing the variables within the system).

(After choosing language and their identification the user initiates DPW.)

S: How may I help you?

U: I would like to create web page that contains information about my contacts, studies and hobbies.

(Personal page is the only type of presentation that allows the requested types of information to be included. Thus, the type of presentation is set to "personal page". Dialogue manager inform the user about inferred piece of information and asked them for the missing mandatory piece of information as well.)

S: You are creating the personal page. Please, enter the name of your presentation.

U: My first page.

(The state of the dialogue is reevaluated. Now, the contact information are expected.)

Ex. 2. Example of a dialogue between a user and BWG system (inference and sub-dialogues).

(The user initiates DPW to choose the type of information.)

S: How may I help you?

U: I would like to include contacts, hobbies and cv into the presentation.

(The system stores the mentioned types. The menu of the chosen types contains the second level but the user does not specify if they have to be used.)

S: OK. Would you like to specify these types of information more precisely?

U: No.

S: OK.

(The state of the dialogue is reevaluated. Now, the contact information are expected.)

Ex. 3. Example of a dialogue between a user and BWG system (choice of the menu items).

4 Conclusions and Future Work

In this paper, we have discussed some methods used in the BWG online application. For visually impaired users, this application is an efficient alternative to web authoring tools. They are encouraged to create and edit fully accessible blogs and personal pages. The efficiency of BWG is enhanced by integrating the module that allows the user to control the application in natural language. Preliminary evaluation results confirm that the alternation between user and system initiative based on users requests is a suitable strategy to increase the system efficiency.

Future work will be dedicated to further extending the functionality in the DPW module in order to minimize the number of misunderstandings (e.g., by involving the machine learning methods). We also intend to analyze the corpus of the prompts and inputs that will be recorded during the additional testing with visually impaired testers.

Acknowledgement

The authors are grateful to the students and staff of the Support Center for Students with Special Needs of Masaryk University for their advice, support and collaboration. This work has been supported by the Grant Agency of Czech Republic under the Grant 201/07/0881.

References

1. Bunt, H.: Context Representation for Dialogue Management. In: Proceedings of the Second international and interdisciplinary Conference on Modeling and Using Context, London, pp. 77–90 (1999)

2. Jönsson, A.: A model for habitable and efficient dialogue management in natural language interaction. In: *Natural Language Engineering*, pp. 103–122 (1997)
3. Griol, D., Hurtado, L.F., Segarra, E., Sanchis, E.: A Dialog Management Methodology Based on Neural Networks and Its Application to Different Domains. In: *Proceedings of the 13th Iberoamerican Congress on Pattern Recognition: Progress in Pattern Recognition, Image Analysis and Applications, Havana, Cuba*, pp. 643–650 (2008)
4. Stent, A., Dowding, J., Gawron, J., Owen Bratt, E., Moore, R.: The CommandTalk spoken dialogue system. In: *ACL 1999*, pp. 183–190 (1999)
5. Litman, D.J., Pan, S.: Designing and evaluating an adaptive spoken dialogue system. In: *User Modeling and User-Adapted Interaction* (2002)
6. Chu-Carroll, J.: MIMIC: An adaptive mixed initiative spoken dialogue system for information queries. In: *Proceedings of the Sixth Conference on Applied Natural Language Processing, Seattle, WA*, pp. 97–104. AAAI Press, Menlo Park (2000)
7. Fügen, C., Holzapfel, H., Waibel, A.: Tight coupling of speech recognition and dialog management - dialog-context grammar weighting for speech recognition. In: *Proceedings of ICSLP 2004, Jeju Island, Korea* (2004)
8. Xu, W., Rudnicky, A.: Language modeling for dialog systems. In: *Proceedings of the ICSLP, Beijing, China* (2000)
9. Visweswariah, K., Printz, H.: Language models conditioned on dialog state. In: *Proceedings of the Eurospeech, Aalborg, Denmark* (2001)
10. Kopeček, I., Bártek, L.: Web Pages for Blind People – Generating Web-Based Presentations by means of Dialogue. In: *Computers Helping People with Special Needs: 10th International Conference, Linz, Austria*, pp. 114–119 (2006)
11. Bártek, L., Plhák, J.: Visually Impaired Users Create Web Pages. In: *11th International Conference on Computers Helping People with Special Needs, Linz, Austria*, pp. 466–473 (2008)
12. Bártek, L., Kopeček, I., Ošlejšek, R.: Setting Layout in Dialogue Generating Web Pages. In: *Matoušek, V., Mautner, P. (eds.) TSD 2007. LNCS (LNAI), vol. 4629*, pp. 613–620. Springer, Heidelberg (2007)
13. Kopeček, I.: Ontology and Knowledge Based Approach to Dialogue Systems. In: *Proceedings of the 2009 International Conference on Internet Computing ICOMP, Las Vegas*, pp. 291–295 (2009)
14. Web Content Accessibility Guidelines (WCAG) 2.0, <http://www.w3.org/TR/WCAG20/>
15. O'Connor, J.: Choosing an accessible CMS (2007), <http://juicystudio.com/article/choosing-an-accessible-cms.php>
16. Kopeček, I., Tiršel, F.: Supporting Visually Impaired People in Image Accessible Exploration and Creation of Personal Presentations. In: *Hlaváčková, D., Horák, A., Osolobě, K., Rychlý, P. (eds.) After a Century of Slavonic Natural Language Processing. Tribun EU 2009*, pp. 135–142 (2009)
17. The Extensible HyperText Markup Language, <http://www.w3.org/TR/xhtml1/>
18. Kopeček, I., Ošlejšek, R.: GATE to Accessibility of Computer Graphics. In: *11th International Conference on Computers Helping People with Special Needs, Linz, Austria*, pp. 295–302 (2008)
19. Kopeček, I., Ošlejšek, R.: Creating Pictures by Dialogue. In: *Computers Helping People with Special Needs: 10th International Conference, Linz, Austria*, pp. 61–68 (2006)
20. Forms in HTML documents, <http://www.w3.org/TR/html401/interact/forms.html>

Design and Development of Spoken Dialog Systems Incorporating Speech Synthesis of Viennese Varieties

Michael Pucher¹, Friedrich Neubarth², and Dietmar Schabus¹

¹ Telecommunications Research Center Vienna (FTW),
Donau City Str. 1, 1220 Vienna, Austria
{pucher,schabus}@ftw.at

² Austrian Research Institute for Artificial Intelligence (OFAI),
Freyung 6/6, 1010 Vienna, Austria
friedrich.neubarth@ofai.at

Abstract. This paper describes our work on the design and development of a spoken dialog system, which uses synthesized speech of various different Viennese varieties. In a previous study we investigated the usefulness of synthesis of varieties. The developed spoken dialog system was especially designed for the different personas that can be realized with multiple varieties. This brings more realistic and fun-to-use spoken dialog systems to the end user and can serve as speech-based user interface for blind users and users with visual impairment. The benefits for this group of users are the increased acceptability and also comprehensibility that comes about when the synthesized speech reflects the user's linguistic and/or social identity.

Keywords: Spoken dialog system, speech synthesis, dialect.

1 Introduction

Speech varieties can be regarded as a primary means for expressing a persons social affiliation and identity. In the context of human computer interaction (HCI) future applications that use speech as their main modality must be capable of reproducing different language varieties. As noted in [2] HCI must shift towards a paradigm where technologies are available to support the production of culture and identity. This is even more important as computers enter everyday life in a more ubiquitous way. Future computers will not only serve as tools or provide information, but will also be monitoring, participating and communicating with users whose experience with such situations and interfaces may be low. Systems that can adapt to the user by displaying a familiar cultural and social identity will definitely increase the acceptance of such systems.

A second aspect, specifically relevant for blind users or users with visual impairment, is that the communicative modality based on spoken language has to serve the needs for both, fast information transmission (fast speech) and comprehensibility. If a user with little experience is confronted with a system that

speaks to her in a language variety that is not familiar to her, comprehensibility is likely to decrease, on a par with acceptability. In the case of a pluricentric language like German, this is obviously the case - at present interfaces using synthetic speech are available, but only generating speech reflecting the standard language spoken in Germany, not the standard spoken in Austria. Non standard varieties are normally not even taken into consideration.

Within the project "Viennese Sociolect and Dialect Synthesis" [1] we developed four synthetic voices, each of them representing a linguistic variety spoken in Vienna. The voice representing Standard Austrian German is already integrated into a web-reader application.

2 Persona Design for Spoken Dialog Systems

2.1 Spoken Dialog Systems

Spoken dialog systems are computer programs that allow a user to interact with a system using speech. Spoken dialog systems are the most advanced form of voice user interfaces (VUI) [3] since they allow for full speech interaction. These systems are composed of three main components: speech synthesis for generating speech output, speech recognition for processing the acoustic input, and dialog management. Equipped with speech synthesis the dialog system is able to transform text available in written form into spoken language. Speech recognition is employed to transform spoken user utterances into written text or words using grammars. The speech recognition component can also be extended by a natural language understanding component. The dialog management component defines the interaction behavior or dialog logic of the dialog system. A standardized definition language for spoken dialog systems is VoiceXML [7], a markup language which is built around the web-based form filling paradigm. The main components of a VoiceXML application are prompts (define what is said), grammars (define what can be said), and forms (define the dialog logic). Figure 1 in section 3 shows a state diagram of our dialog system. All green blocks are defined within one VoiceXML page using four different forms. One form for the selection of the restaurant and a separate form for each restaurant. Within a restaurant form it is possible to make a phone call to the selected restaurant.

2.2 Persona Design

The personality or persona of a spoken dialog system must be considered, since there is no such thing as a voice user interface with no personality [3]. The perception of sociolectal and dialectal varieties influences our evaluation of a speaker's attributes like competence, intelligence, and friendliness. The persona can be defined as a standardized mental image of a personality or character that users infer from the applications voice and language choices [3]. Speech synthesis is an essential part of a spoken dialog system's persona. In a previous study [4] we found the following positive and negative attributes that are associated with sociolect / dialect voices. It should be mentioned that this study was carried

Table 1. Positive and negative attributes of sociolect / dialect voices

Positive (application)	Negative (application)
fun (game)	not respectable (banking)
optional (navigation)	not formal (banking)
personal (navigation)	not neutral (administration)
regional (taxi)	not intelligible (flight domain)
democratic (administration)	non-native spk. (flight domain)
persona (taxi)	lack of precision (health)
tourist (district info)	lack of prestige (banking)
lack of trust (administration)	lack of competence (health)

out with special focus on varieties within Austria, but it can be assumed that the results will match users' judgments about many dialect voices that stand in opposition to an officially recognized standard variety.

3 Spoken Dialog System

Within our research project on synthesis of Viennese varieties [1] we developed four synthetic voices that represent representative points in a 3-dimensional space of language varieties, which is defined through age, gender, and education [6].

1. Austrian German standard (35–50, male, +)
2. Viennese dialect (45–60, male, –)
3. Viennese youth language (15–25, female, +/-)
4. Viennese standard (55–70, female, +)

On the basis of this selection of prototypes, we created four different personas within our spoken dialog system. In the evaluation of possible scenarios [4] we found that a restaurant information system is well suited for Viennese dialect synthesis. The mapping of positive / negative properties to standard / dialect served as design guideline for the dialog design. The speaker of the Austrian standard variety (1) was used as a moderator, also providing some help functionality. Every other speaker recommends a different type of restaurant, which is associated with the speaker's sociolect.

Figure 1 shows the VoiceXML [7] dialog flow diagram for our regionalized and localized restaurant guide. Speaker specific prompts were used for the dialog

Table 2. Sociolect / restaurant type association

Speaker sociolect	Restaurant type
2. Viennese dialect	Viennese cooking
3. Viennese youth language	Low prices / cool places
4. Viennese standard	Luxury restaurants

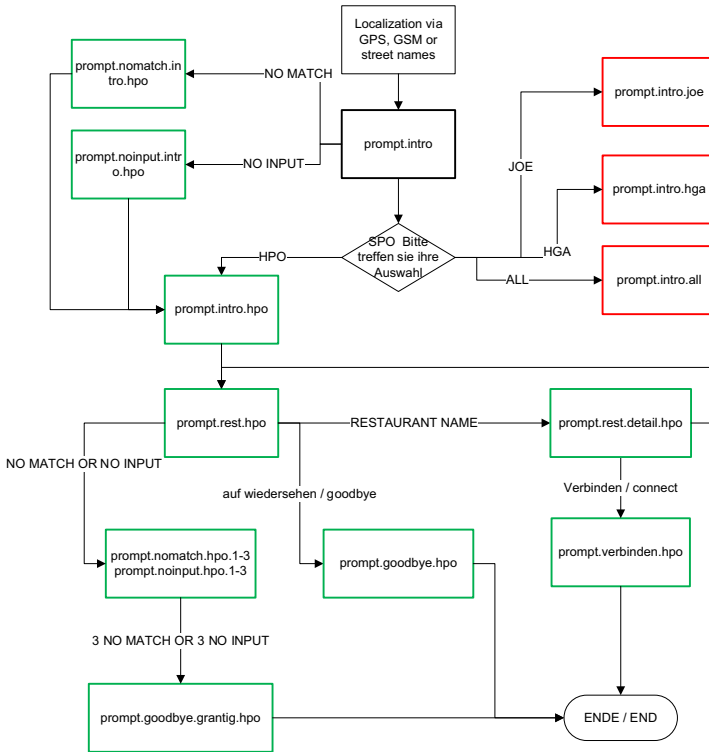


Fig. 1. VoiceXML dialog flow

and no match / no input strategies. One additional dialog with all speakers and interactions between them was also realized. On our project webpage [1] it is possible to try out all four synthetic voices as well as the dialog system.

4 Blending between Standard and Dialect

To enable applications that use continuous varieties between standard and dialect we have investigated interpolation methods that can realize such in-between varieties [5]. The blending between an Austrian German and Viennese utterance is shown in Figure 2. This can be realized using the parametric synthesis method of Hidden Markov Model (HMM) based speech synthesis. With this method it is possible to interpolate models of different varieties and thereby create in-between varieties. By interpolating the duration model with a zero duration model we can also model units that are present in one variety but not in the other (such as vowel deletion/epenthesis) as shown in Figure 2. To model context-dependency of units a broad linguistic context is taken into account consisting of linguistic features on different linguistic levels.

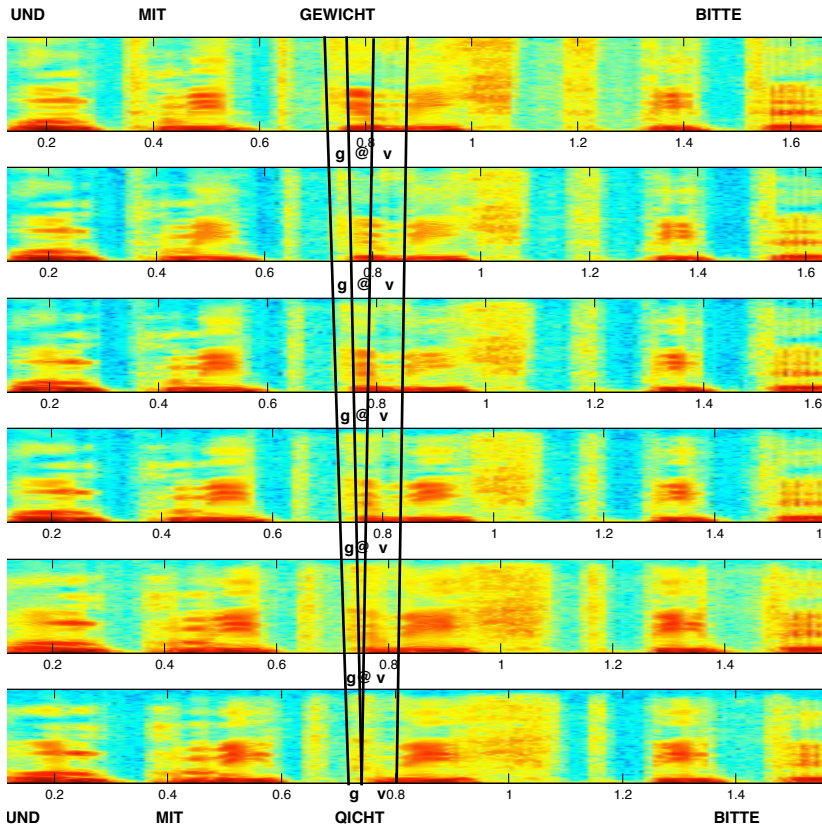


Fig. 2. Blending between German “Und mit Gewicht bitte” and Viennese “Und mit Qicht bitte”

These blending methods could be used in a dialect / sociolect adaptive spoken dialog system. By using dialect recognition there could be even a matching of the speaker’s dialect and the dialect of the spoken dialog system. In this way it would be possible to gradually adapt to the user’s dialect. The variety of the dialog system could be chosen to be similar [8] or different from the speaker’s variety. [8] shows that users prefer accents that are similar to their own accent. This, however, is not necessarily true for dialects or sociolects.

5 Conclusion

We discussed the importance of regionalized adaptive user interfaces and showed the design of a spoken dialog system that incorporates multiple varieties which represent the sociolect space of Vienna. In future work we want to develop additional spoken dialog systems that are suitable for sociolect / dialect synthesis.

Furthermore we also want to investigate less known varieties like Turkish-Viennese, which enable new personas.

Acknowledgements

The project “Viennese Sociolect and Dialect Synthesis” was funded by the Vienna Science and Technology Fund (WWTF). The Telecommunications Research Center Vienna (FTW) is supported by the Austrian Government and the City of Vienna within the competence center program COMET. OFAI is supported by the Austrian Federal Ministry for Transport, Innovation, and Technology and by the Austrian Federal Ministry for Science and Research.

References

1. Viennese Sociolect and Dialect Synthesis, <http://dialect-tts.ftw.at>
2. Cassell, J.: Social Practice: Becoming Enculturated in Human-Computer Interaction. In: Stephanidis, C. (ed.) Universal Access in HCI (UAHCI), HCI 2009. LNCS, vol. 5616, pp. 303–313. Springer, Heidelberg (2009)
3. Cohen, M.H., Giangola, J.P., Balogh, J.: Voice user interface design. Addison-Wesley, Reading (2004)
4. Pucher, M., Schuchmann, G., Fröhlich, P.: Regionalized Text-to-Speech Systems: Persona Design and Application Scenarios. In: Esposito, A., Hussain, A., Marinaro, M., Martone, R. (eds.) Multimodal Signals: Cognitive and Algorithmic Issues. LNCS (LNAI), vol. 5398, pp. 216–222. Springer, Heidelberg (2009)
5. Pucher, M., Schabus, D., Yamagishi, J., Neubarth, F., Strom, V.: Modeling and Interpolation of Austrian German and Viennese Dialect in HMM-based Speech Synthesis. *Speech Communication* 52(2), 164–179 (2010)
6. Moosmüller, S.: Soziophonologische Variation im gegenwärtigen Wiener Deutsch. Franz Steiner Verlag, Stuttgart (1987)
7. VoiceXML 2.0 recommendation, <http://www.w3.org/TR/voicexml20/>
8. Dahlbäck, N., Wang, Q., Nass, C., Alwin, J.: Similarity is more important than expertise: Accent effects in speech interfaces. In: Proc. SIGCHI conference on human factors in computing systems, pp. 1553–1556 (2007)

Methods for the Visually Impaired to Retrieve Audio Information Efficiently

Atsushi Imai^{1,3}, Nobumasa Seiyama¹, Tohru Takagi², and Tohru Ifukube³

¹ NHK Engineering Services, Inc.,

1-10-11 Kinuta, Setagaya-ku, Tokyo, 157-8540, Japan

² NHK (Nippon Hoso Kyokai; Japan Broadcasting Corp.)

Science and Technology Research Laboratories,

1-10-11 Kinuta, Setagaya-ku, Tokyo, 157-8510, Japan

{imai.a-dy, seiyama.n-ek, takagi.t-fo}@nhk.or.jp

³ Resarch Center for Advanced Science and Technology, The University of Tokyo,

4-6-1 Komaba, Meguro-ku, Tokyo, 153-8904, Japan

{imai, ifukube}@rcast.u-tokyo.ac.jp

Abstract. Visually impaired persons are able to get large amounts of information through sound, especially speech, but because sound information is time-sequential it is difficult to gain an overall understanding of content in a short time, presenting inevitable issues. This paper described a solution we have proposed a framework for fast listening methods (e.g. for 3 times normal speed or more over) that are the equivalent of visual “skimming” of textual reading materials. Currently we have completed a basic study and preliminary tests indicate positive results.

Keywords: blind user, speech rate conversion, adaptive speech rate conversion, fast listening, recording book, scanning, skimming.

1 Introduction

When visually able people read a magazine or news paper, they do so at various levels of detail, from “close reading”, paying attention to each letter and phrase, to “skimming” for the overall meaning. The visually impaired persons also have a similar need when obtaining information from audio sources such as recorded books or PC screen reader and so on, but currently there is no equivalent to “scanning” for playback of audio.

Thus, we have developed a method for visually impaired persons to obtain information from audio sources efficiently that is equivalent to skimming of text material, and we report on this method below. This paper described a solution we have proposed a framework for fast listening methods (e.g. for 3 times normal speed or more over). We are aiming to exceed the maximum replay rate of the function that is mounted on a current prevailing device.

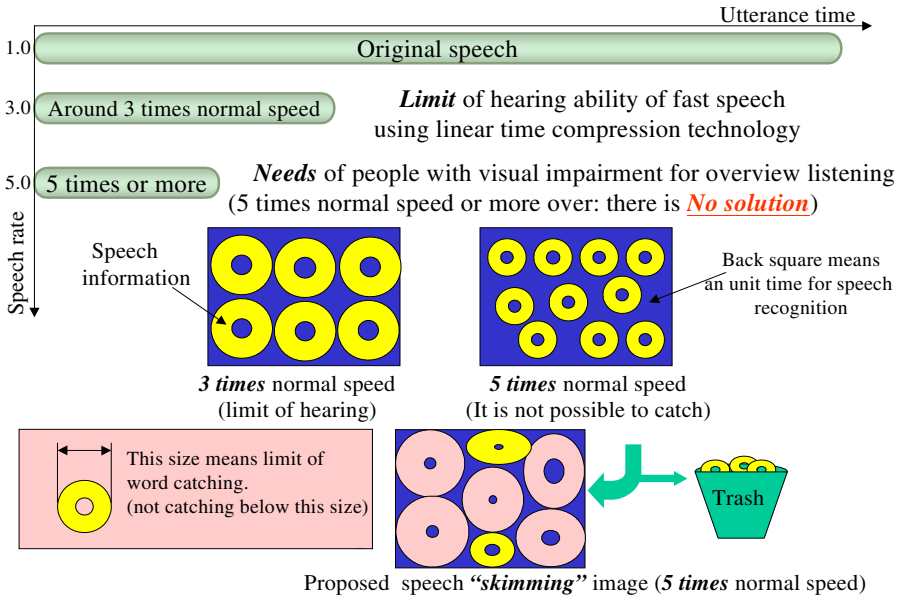


Fig. 1. Outline of proposed method for ultra rapid listening

2 Background

Earlier research has shown that listening comprehension when the playback speed of audio is increased uniformly is limited to approximately three times the original speed [1]. There have also been reports of adjusting playback speed adaptively rather than uniformly, which have achieved speed up to five times the original speed [2]. However, many visually impaired persons have indicated a need to get information even more quickly than this [3][4]. A certain amount of increase in the limits of hearing comprehension could be possible by optimizing these earlier methods, but dramatic increases in speed, to meet the needs of our subjects, are not likely. Thus, we have devised a method to emulate “skimming” in which, rather than attempting to play back all audio information, we proactively skip sections of less important information and play back the remaining information at high speed in a manner as easy to comprehend as possible. Figure 1 illustrates this concept.

3 Algorithms

The basic concept is to retain parts that are likely to have a significant effect on hearing comprehension, deleting the rest. For example, within a continuous utterance, if both the power and the pitch of the voice are relatively low in a particular section, that section could be removed. This type of section would be difficult to understand even when played back at normal speed, so we can proactively delete it.

Since operation would not be practical unless it was fully automatic, specific acoustical feature that can be obtained by audio analysis must be used as clues for this process. For the example mentioned earlier, we can create a rule of speech rate control for fast listening that observes time fluctuations in the power and pitch of the speech signal and deletes sections when they fall below a certain threshold. In addition, existed speech signal that is not deleted is also converted the speech rate adaptively along with the temporal change of pitch and power. Speech rate controls are done by our original method [5]. Playback speed can be controlled using arbitrary functions, and in particular, fluctuations in pitch or power can be used to control speed. The actual physical quantities we have analyzed are pitch, power, and duration and locations of continuous sections of speech signals.

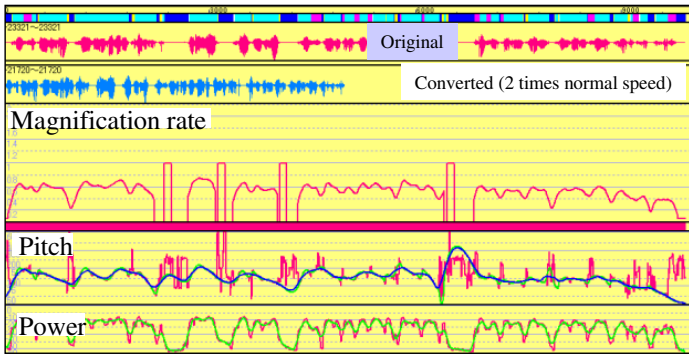


Fig. 2. Example of adaptive speech rate (=magnification rate) setting

3.1 Trial Method for Fast Listening

Before making optimum conditions of fast listening method, we conducted preliminary study and evaluation with one visually impaired subject and six sighted subjects including the author to evaluate hearing comprehension and identify other potential problems or discomforts using trial converted speeches described below.

We selected about 40-second segments from literature in the DAISY recorded library (in Japanese) as our listening material. The test audio segments were prepared under the following conditions:

- (Control) Uniform time compression for 2 to 4 times normal speed
- (Proposal) Sections of silence greater than 200 ms are deleted, retaining only 100 ms. Up to 250 ms of sound before the silent section is also deleted. The 100 ms sections of silence are retained, and the remaining speech is converted according to a function that places a 50% weighting on each of pitch (=Pi[Hz]) and power (=Po[dB]).
- Finally, the overall playback time is adjusted such that the total playback time is the same as that for 2 to 4 times normal speed playback. Speech rate (=magnification rate) is calculated by equation (1):

$$f(P_o, P_i) = 0.1 + (\sqrt{(2 \times (55 + P_o))} \times \sqrt{P_i}) / 120 \quad (1)$$

The calculated magnification curve always becomes 0.1 when power is very small (-55dB) and when both values are close to the mean value, it becomes around 1.0. If either of value is larger, the magnification is also relatively larger. Figure 2 shows an example of speech rate setting.

The playback time for each pair of the converted speech segments above (Control and Proposal) was completely same. Preliminary tests indicated that the methods should be quite effective, and we intend to conduct full-scale tests, and to increase the variety of conditions used to create speech samples.

4 Evaluation

While the algorithm was customized in Japanese, it was applied to English with slight adjustment. We selected the experimental listening material from literature in the DAISY recorded library (in English), CNN news program and TOEFL tests. Either speech segment is about 30ms. The experiment evaluated about the point of view “which is easier to catch the meaning” of a pair of Control/Proposal speech segment.

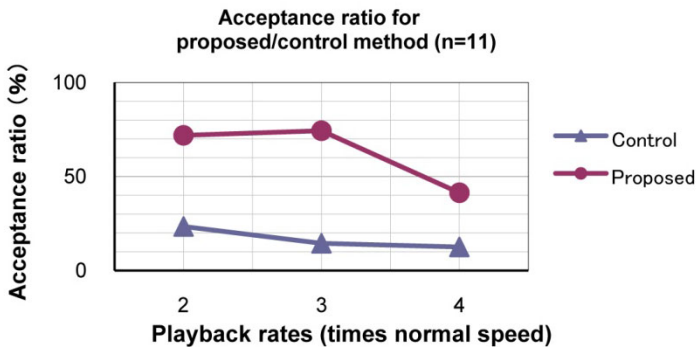


Fig. 3. Result of acceptance ratio for proposed conversion and control (=conventional) conversion at 2 to 4 times normal speed

The subjects selected desirable one. The playback rate of 2 and 3 times normal speed were evaluated 10 times each evaluation point, and 4 times normal speed was 4 times also. The experimental materials were presented in consideration of the order effect.

According to the evaluation by 11 native speakers of English and also visually impaired subject, the speech produced using the conventional method was mostly incomprehensible at 4 times normal speed, while that from the proposed methods was comparatively comprehensible. However, the result was profoundly affected by the order of presenting the control speech and the proposed one. Figure 3 shows the general result. Figure 4 shows results of each order of presentation.

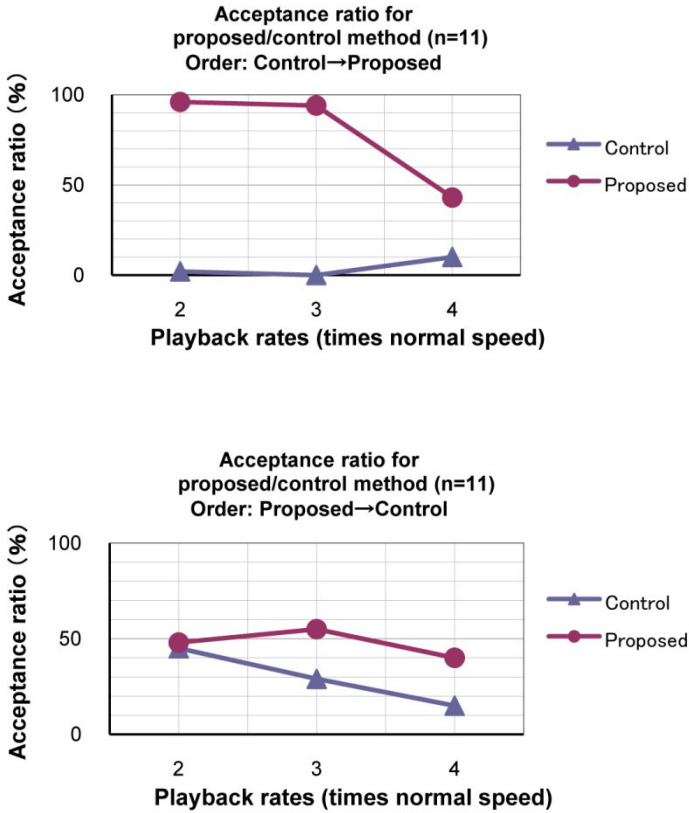


Fig. 4. Result of acceptance ratio for proposed conversion and control (=conventional) conversion at 2 to 4 times normal speed by each order of presentation (Control/Proposed)

It was a similar tendency in each presentation order at 4 times normal speed but 2 and 3 times normal speed showed different trends. As a result, it is thought that there was significant effectiveness in at least 4 times speed. That means effectiveness can be confirmed by a more high-speed reproduction.

5 Summary and Future Work

In this research we have proposed a framework for listening methods that are the equivalent of visual “skimming” of textual reading materials. Currently we have completed a basic study and preliminary tests indicate positive results, so we plan to proceed with more detailed research in the future. We consider that our proposed method depends on the ability that the person predicts the speech context instantaneously. So, we plan to examine the relation to the context effect more in detail.

References

1. Asakawa, C., et al.: The optimal and maximum listening rates in presenting speech information to the blind. *Journal of Human Interface Society* 7(1), 105–111 (2005) (in Japanese)
2. Imai, A., et al.: Application of Speech Rate Conversion Technology to Video Editing - Allows up to 5 times normal speed playback while maintaining speech intelligibility. In: 20th International Conference on Audio Engineering Society, pp. 96–101 (October 2001)
3. Torihara, S.: An “Oblique” Listening Method for the Blind and Visually Impaired. In: The Sixteenth Annual Conference on Technology and Persons with Disabilities, CSUN 2001, California State University, Northridge (March 2001)
4. Torihara, S., Nakamura, M., Ueda, N., Wada, T., Ishizaki, S.: English “Oblique” Listening System – Rapid Listening System for the Blind and Visually Impaired, and Its Evaluation. In: Miesenberger, K., Klaus, J., Zagler, W.L., Karshmer, A.I. (eds.) ICCHP 2006. LNCS, vol. 4061, pp. 53–60. Springer, Heidelberg (2006)
5. Imai, A., et al.: An Adaptive Speech-Rate Conversion Method for News Programs without Accumulating Time Delay. *IEICE Transactions A* J83-A(8), 935–945 (2000) (in Japanese)

Binocular Vision Impairments Therapy Supported by Contactless Eye-Gaze Tracking System

Łukasz Kosikowski and Andrzej Czyżewski

Faculty of Electronics, Telecommunications and Informatics,
Gdansk University of Technology, Gdańsk, Poland
{kosiq, andcz}@sound.eti.pg.gda.pl

Abstract. Binocular vision impairments often result in partial or total loss of stereoscopic vision. The lack of binocular vision is a serious vision impairment that deserves more attention. Very important result of the binocular vision impairments is a binocular depth perception. This paper describes also a concept of a measurement and therapy system for the binocular vision impairments by using eye-gaze tracking system.

Keywords: Amblyopia, lazy-eye syndrome, eye-gaze tracking system.

1 Introduction

The loss of binocular depth perception robs a person the possibility of being for example a dentist and dramatically decreases quality of life. Parents and patients need to be informed about early detection of these conditions as well as the full range of treatment options. One type of binocular vision impairment is amblyopia or so called lazy eye syndrome affects approximately 3% of people. Parents or teacher do not easily detect some binocular vision impairments [1].

There are several systems on the market, which allows treating or measuring lazy-eye syndrome by using PC-like computers [2], [3], [4]. Some of them are based on virtual reality [2]. Most of the systems are focused only on therapy of amblyopia without measuring any parameters or vice-versa they are based on measurements, only. The proposed approach provides a complete measurement and therapy system for detecting and treating binocular vision impairments - especially strabismus (small and middle range) and lazy-eye syndrome [5]. To support objective measurements it was extended by an eye-gaze tracking system (Fig. 1). Eye-gaze tracking systems are successfully used in the eye examination [6]. The contactless eye-gaze tracking system was developed at the Multimedia System Department, Gdansk University of Technology and was successfully applied in other applications, requiring precise measurements [7], [8]. The eye-tracking system is characterized by the lack of any additional physical equipment mounted on the user's body and the fixation point detection process is hardware-supported by infrared diodes [7].

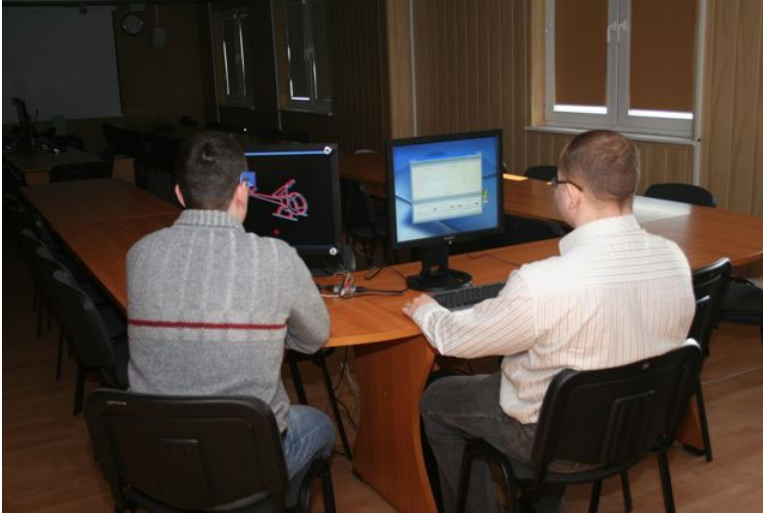


Fig. 1. Illustration of the prepared system

2 System Description

The described system allows for measuring a set of parameters. Data are obtained in center points (so-called control points) of 9 areas of equal size on which patient directs his eyes holding his head still [5]. The fixation point is controlled by the eye-gaze tracking system. The computer screen was divided in such a way that the angle of seeing and range of measured parameters are maximized.

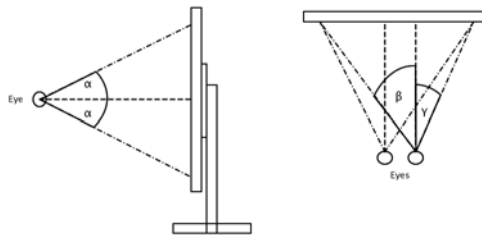


Fig. 2. Eye-gaze measurement angles

Because of the limitations of the eye-gaze tracking system and the maximization of range of the measured parameters, the measurements are acquired in 9 control points. Measurements are conducted for set of angles: 0° , -11° and $+11^\circ$ in vertical plane, and 0° , -15° , $+15^\circ$ in horizontal plane (Fig. 2). The angle $\alpha=11^\circ$, and the $\beta+\gamma=2\times 15^\circ$.

The examined person has to modify available transformation parameters of image (vertical and horizontal position, scale rotation) to finally see one image (two overlapping images). The person who wears glasses should be wearing them during the test. Modification of the height of the image allows eliminating the impact of

anisometropia (the condition in which both eyes have unequal refractive power), modification of the width of the image allows eliminating the impact of regular astigmatism, that is relevant to determining the proportion of width to height, Modification of the vertical position parameter allows compensating hypertropia or hypotropia (eye that turns upwards or downwards respectively) image distortion results, modification of the horizontal position parameter allows compensating the squint angle. In addition, this parameter depends on the distance of eyes from the computer monitor, modification of the rotation parameters between -45 and $+45$ degrees allows compensating oblique squint (strabismus obliquus) image distortion results [5].

The main goal of the system was analyzing the change depending on the transformation of images as a function of distance from the center of the screen. The set of 9 vectors of measurement parameters allows calculating the image transformation parameters at any point of the screen. The images dedicated to the left and to the right eye are transformed by using interpolated parameters in such a way that the user looking at the screen sees one image (two overlapped images). As a result of synthesis of transformed images in the brain the "baseline" between the eyes is produced, rendering a single integrated image [5]. Regular therapy allows creating the stereovision.

In the system many different exercises are implemented. The system allows the user to perform therapy, using a set of vectors obtained from the examination process. Images displayed on the monitor are modified according to those results. During the therapy a user has to focus eyes on the screen. The fixation point position is detected, and the presented images are geometrically transformed using the results from measurement process. Because the measurement process is discrete, the interpolation and extrapolation algorithms of the results are required. When user looks at the screen between coordinates of the measurement points, the new transform factors are calculated.

The distance equals to 19" between human eyes and the LCD display is controlled by the gaze-tracking system. That system precisely analyzes the eye gazes and head position, and allows for taking measurements with a high precision. The person who wears glasses should be wearing them during the test and the therapy. The data access module provides mechanisms for the storage and access to data. The database stores information about users (patients) and all information about examinations and therapies, which allows for tracking progress of the therapy [5].

3 Conclusions

It should be remembered that most of the target users of this system are school children, so it is of utmost importance that difficulties in learning (including amblyopia) should be detected and eliminated at the earliest stage of their education.

We are planning to change the image display method. Currently, the system exploits the anaglyph method, thus we plan to transform this method to a stereo image system employing 3D LCD monitors and polarization glasses. This modification would allow for increasing image quality and accuracy of the measurement system.

Acknowledgments

Research funded within the project No. POIG.01.03.01-22-017/08, entitled "Elaboration of a series of multimodal interfaces and their implementation to educational,

medical, security and industrial applications". The project is subsidized by the European regional development fund of the and by the Polish State budget".

References

1. Optometrists network, <http://www.children-special-needs.org>
2. Qiu, F., Liu, Y., Wang, L., Yu, L.: Interactive binocular amblyopia treatment system with full-field vision based on virtual reality. In: Int. Conf. Bioinformatics and Biomedical Engineering, pp. 1257–1260 (2007)
3. Qiu, F., Wang, L., Liu, Y., Li, H.: Research on Interactive System of Amblyopia Treatment, In: The 1st International Conference on Bioinformatics and Biomedical Engineering, ICBBE 2007, pp. 1177–1180 (2007)
4. Van Eenwyk, J., Agah, A., Cibis, G.W.: Automated human vision assessment using computer vision and artificial intelligence. In: IEEE International Conference on System of Systems Engineering, SoSE 2008, pp. 1–6 (2008)
5. Kosikowski, Ł., Czyżewski, A.: Computer Based System for Strabismus and Amblyopia Therapy. In: 2nd International Symposium on Multimedia – Applications and Processing, pp. 407–410. Mrągowo (2009)
6. Augustyniak, P., Tadeusiewicz, R.: Ubiquitous Cardiology: Emerging Wireless Telemedical Applications. Medical Information Science Reference, Hershey, New York (2009)
7. Czyżewski, A., Kunka, B., Kurkowski, M., Branchat, R.: Comparison of developed gaze point estimation methods. In: Signal Processing: Algorithms, Architectures, Arrangements, and Applications, SPA IEEE 2008: New Trends in Audio and Video, pp. 133–136. Poznań (2008)
8. Kunka, B., Czyżewski, A., Kostek, B.: Concentration tests: An application of gaze tracker to concentration exercises. In: Proc. 1st International Conference on Computer Supported Education CSEDU 2009, Lizbona, p. 66 (2009)

Realization of Direct Manipulation for Visually Impaired on Touch Panel Interface

Tatuso Nishizawa^{1,4}, Daisuke Nagasaka², Hiroshi Kodama³, Masami Hashimoto³,
Kazunori Itoh³, and Seiji Miyaoka⁴

¹ Interdisciplinary Graduate School of Science and Technology, Shinshu University, 4-17-1
Wakasato Nagano-Shi Nagano 386-8553 Japan

tatsu12@attglobal.net

² Graduate School of Science and Technology, Shinshu University, 4-17-1 Wakasato
Nagano-Shi Nagano 386-8553 Japan

³ Faculty of Engineering, Shinshu University, 4-17-1 Wakasato
Nagano-Shi Nagano 386-8553 Japan

{t09a542,t065040,hasimoto,itoth}@cs.shinshu-u.ac.jp

⁴ Shinano Kenshi Co., Ltd. 1078 Kamimaruko Nagano 386-0498 Japan
seiji_miyaoka@skcj.co.jp

Abstract. Touch panel interface is expanding its application rapidly, with the great flexibility and GUI that allows direct manipulation, and allows easy interaction for sighted people. Various approaches have been tried to achieve accessible touch panel for the visually impaired by combination of tactile sense and audio feedback. In this paper, we report the real benefit of direct manipulation of touch panel for the visually impaired by typical finger gestures such as finger flicks and taps only. Implementations of DAISY digital talking book navigation, by up/down and left/right finger flicks operations were compared with conventional up/down and left/right button operations. As a result, finger flicks achieved the same as or faster in regards to operation.

Keywords: touch panel, visually impaired, finger gesture, mobile phone, DAISY.

1 Introduction

Many recent mobile devices such as the iPhone integrate the touch panel to interact with the device by finger gestures. However this interaction significantly relies on visual feedback, which is inaccessible for the visually impaired.

Typical approaches to improve touch panel accessibility for the visually impaired are taken by the combination of tactile sense and audio feedback. Yatani et al. introduced SemFeel[1]. Another example is screen-reading capability equipped with the iPhone3GS operating system, called “Voiceover”[2]. Also, by utilizing the multi-touch interface feature on the iPhone, Kane et al. studied a gesture called Slide Rule [3]. Their study showed that Slide Rule was faster than a button-based interface and it also showed more errors in Slide Rule by participants. And mBN(Mobile Blind

Navigation)[4] shown the usefulness of dragging the pointer from one corner to another at mobile display.

However these approaches are targeting knowledgeable people as a typical user, we think touch panel will become much more popular in the near future, so that coverage of ages and skills should be expanding, since the benefit of touch screen increases. At this point, we should focus feeling of directness.

Hutchins indicates two distinct aspects of the feeling of directness [5].

1) Notion of the distance between one's thoughts and the physical requirements of the system under use.

2) In a system built on the model-world metaphor, the interface is itself a world where the user can act, and which changes state in response to user actions.

Typical touch panel application shows selection branches simultaneously on the screen. For the visually impaired, it is very hard to recognize this parallel world without assistance of tactile information, and they favor recognition through audio information. However audio information can only be provided sequentially, thus this distance between parallel and serial world causes a big distance at direct manipulation for the visually impaired.

Based on this preliminary study, we chose DAISY [6] digital talking book playback system as implementation target to examine the potential of touch panel interface for the visually impaired. Thus audio playback needs the sequential operations. In this paper, we propose simple finger gestures by typical touch panel gestures without any additional tactile assistance. The performance was compared with conventional push button operation.

2 Basic Experiment

As a primary experiment, we gathered typical practice data of finger gestures to determine adequate assignments for DAISY playback operation. As a typical finger operation, we chose taps and flicks, since they are easy to perform.

With five visually impaired participants from age 20 to 70, Four of them are totally blind, and one of them had limited sight, three of them use a PC regularly, and two of them have little or no experience with a PC, finger gestures with tap (single and double), and flick (up/down and left/right) were measured and compared with sighted people age 20 to 30. All of them had experience with a DAISY player.

2.1 Evaluation Environments

We implemented logging software on the Apple iPhone3GS, that measures touch and release locations with time stamps. The iPhone has 320 pixels horizontally and 480 pixels vertically, so that locations are measured and stored by this pixel, and time is measured with millisecond accuracy. Also, this logging software has tutorial function with audio guidance that gives the instructions to the participants. The sample of logging data of single tap operation is as follows. [Start] X=79, Y=255, t=5.411, [End] X=79, Y=253, time=5.469, where X shows horizontal pixel location and Y shows vertical one, also time shows the timer counter value with mille second accuracy.

2.1.1 Adaptation of the Case Cover

At the beginning, we conducted interviews with five participants by touch and feel the iPhone voiceover, since they never had experience with touch panel interface before. Question is simply asking the difficulty of usability. Then, all of them complained about the difficulty of recognition of the active area of touch panel, since the surface of the iPhone was covered with flat grass. We tried several kinds of commercially available case covers, then found the rubber case cover that clearly indicated the active area of touch panel by tactually. Figure 1 left shows the rubber case cover and original iPhone. Figure 1 right shows the typical handling of iPhone with rubber case.

Rubber case cover and original iPhone



Handling of iPhone with rubber case



Fig. 1. Rubber case cover and the iPhone

2.1.2 Instruction of Tap and Flick Operation

We instructed the participant to grasp the iPhone by left hand, and then use the index finger of right hand for gesturer operation that is also shown in Fig.1 right. Instruction of each evaluation is as follows.

Tap operation: Three times for each position from center, upper left, upper right, lower left, lower right. First round single tap, and then second round double taps.

Flick operation: Three times for each combination of different 90 degrees direction, right right up, right right down, left left up, left left down, up up right, up up left, down down right, down down left, three times each.

2.2 Evaluation Results

Accuracy of tap position: Fig 2 shows the locations of tap operation by visually impaired that compare with sighted people. The visually impaired have a wider spreading out of tap locations and large differences among individuals. At Fig.2 one cell containing 80 pixels, it reflects the screen size of the iPhone. (320x480 pixels)

Interval of double taps: Fig 3 shows the time interval of the double taps by visually impaired when compared with sighted people. Differences among individuals were observed, but no difference between the visually impaired and the sighted people.

Flick operation: Fig.4 shows the length and direction of flick operations by visually impaired when compared with sighted people. At Fig.4 one cell containing the 80 pixels each, starting point of flick operation was normalised at the center of each graph. So that the range of the graph expanded to right and left direction as 320 pixels maximum each and up and down direction as 480 pixels maximum each.

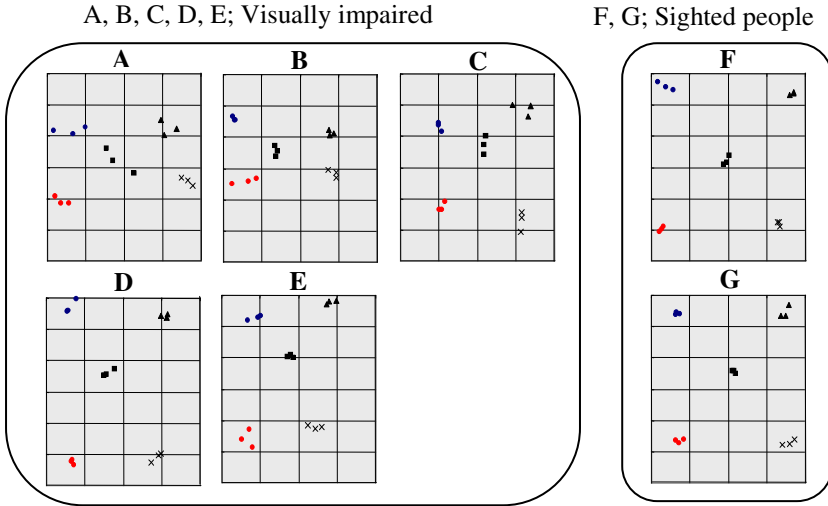


Fig. 2. Distributions of locations on screen with tap operation (80 pixels per cell)

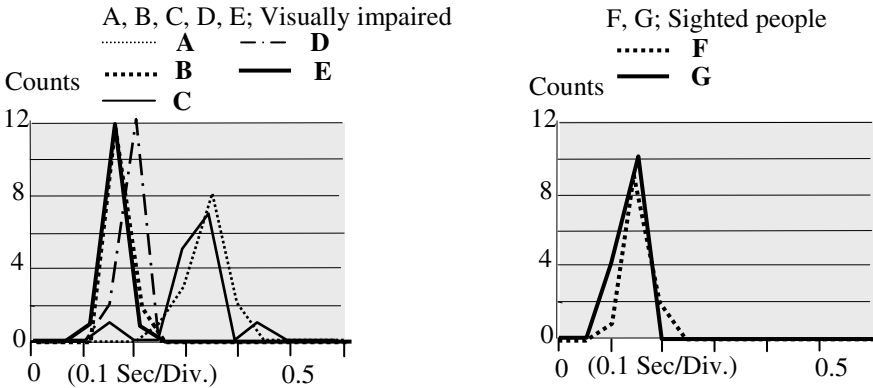


Fig. 3. Distributions of interval of double taps

Figure 5 shows histogram of flick direction, by applying thresholds as plus minus 45 degrees with 0 degree for right, 90 degrees for up, 180 degrees for left and 270 degrees for down, flick direction can discriminate as right, up, left, down. Wider distribution exists at upper direction with average 92 degrees and SD 12.7 degrees.

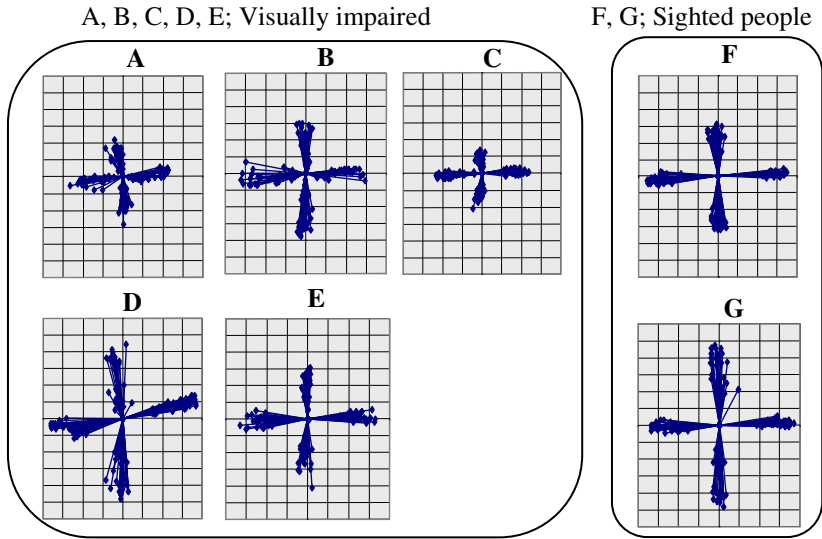


Fig. 4. Distributions of flick length and direction (80 pixels per cell)

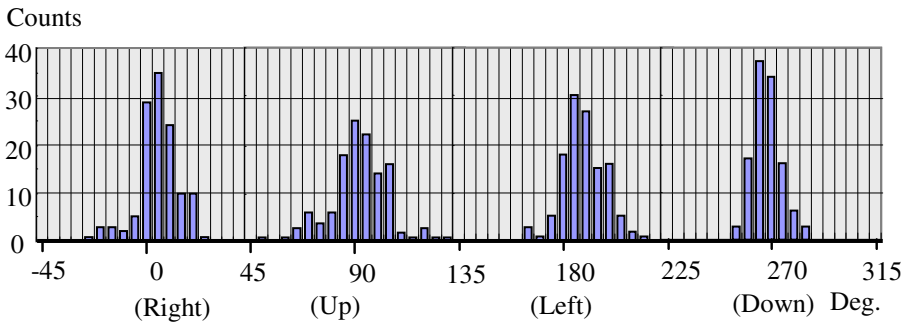


Fig. 5. Histograms of flick direction by visually impaired

2.3 Summary of Basic Experiment

Application software that responds to the accuracy of tap position is not adequate for the visually impaired, however, the difference between single and double tap is clear, so that single tap may apply as the status indication, double taps may apply as confirmation or execution of the task. On the other hand, flick direction is very clear, thus it is suitable for up/down and left/right indication by flick operation.

3 Comparison of Flick Operation and Button Operation

As an actual performance comparison, we chose the navigation of DAISY digital talking book. DAISY was introduced as the standard of digital talking book services

for the visually impaired. In this standard hierarchy structure of the book is represented from h1 to h6 heading tags in NCC.html (Navigation Control Center that is represented as table of contents for navigation). To perform a task of navigation for the specific part of a book, that requires the change of the heading level from h1 to h6 by indication of up/down direction and selects the previous or next heading by indication of left/right direction. Thus, we chose DAISY navigation as implementation target, and these operations have great similarity with menu selection system.

3.1 Evaluation Environments

Figure 6 shows the test under devices and operation assignments for each device. The (1) iPhone utilizes flick and tap gestures, and (2) PTP1 utilizes five buttons for basic navigation. The PTP1 is a portable DAISY player from Shinano Kenshi. [7].

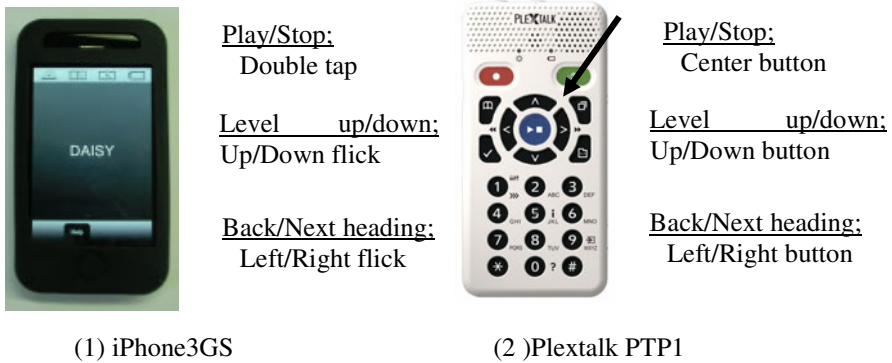


Fig. 6. Under test devices

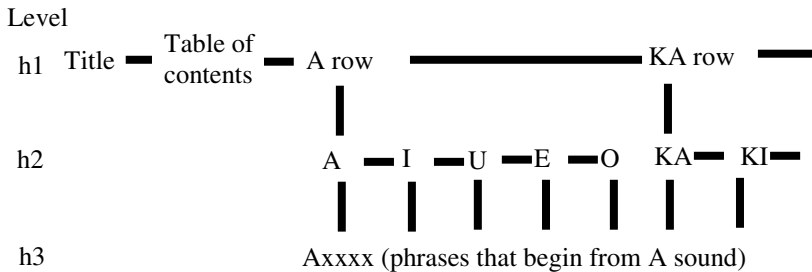


Fig. 7. Under test DAISY book “Dictionary of Beautiful Japanese phrases”

Figure 7 shows the test under book of “Dictionary of Beautiful Japanese phrases” with 3 levels (h1, h2 and h3). The level 1 (h1) is highest level that contains the Japanese kana groupings, followed by table of contents. The level 2 (h2) is middle level that contains each kana character that belongs to the kana grouping. The level 3 (h3) is lowest level that contains the Japanese phrases that begin with level 2 kana character.

3.2 Evaluation Specification

Three tasks were given to the participants, each of the tasks indicate three specific phrases at level 3. The indication of phrases is given sequentially, once the navigation to the specific phrase is performed. Performance was measured by total required time and flick or button operation counts.

3.3 Evaluation Results

Figure 8 shows same or better performance for finger flick operations compared with conventional push button operations. All participants have had experience with exiting DAISY playback systems that utilize push buttons, and for the training of finger flick, it needed just 30 minutes trial that include the basic experience measurements.

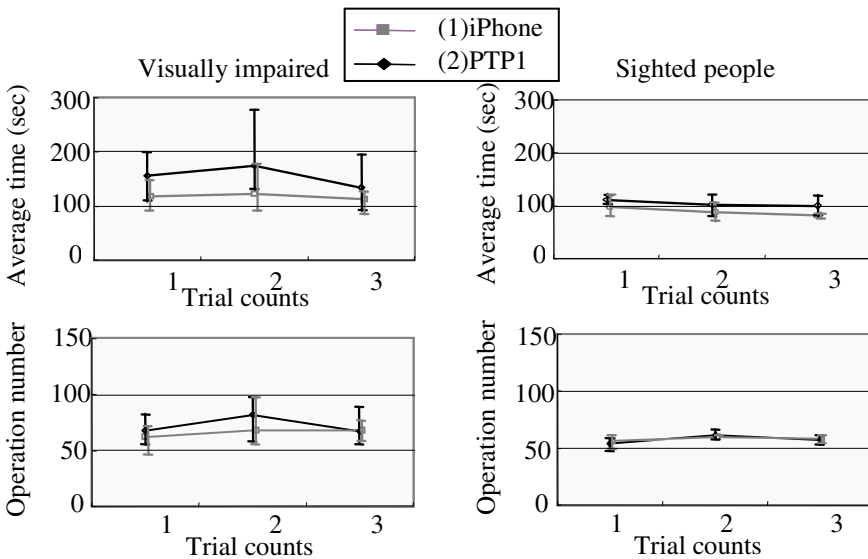


Fig. 8. Performance comparisons between each device to achieve the tasks

4 Conclusion

We introduced direct manipulation of touch panel by finger gestures, which achieved same as or faster than button operations for the navigation of DAISY digital talking book. Positive feedback has highlighted for the simplicity compared with finding the buttons, and negative feedback has highlighted for the unexpected response from the iPhone that was caused by unintentional touch to the panel. We need further study and improvements of the behaviors of touch panel interaction, especially for the unintentional touch.

In addition, as a pilot study, we observed handwriting Japanese character recognition on touch panel by two visually impaired who lost their sight in the middle of their

life, by 20 Japanese kana-characters recognition, they achieved over 90% accuracy, and easily recovered by additional try, and they mentioned that it was very straight forward and useful method as text input for the mobile devices. We believe by utilizing those flexibilities of touch panel interface, finger gestures will be widely accepted and easy to use as a direct manipulation interface for the visually impaired.

References

1. Yatani, K., Truong, K.N.: UIST 2009: Proceedings of the 22nd annual ACM symposium on User interface software and technology (October 2009)
2. iPhone User Guide For iPhone OS 3.1 Software, ch. 25 Accessibility Features 183-189 Apple Inc. (2009)
3. Kane, S.K., Bigham, J.P., Wobbrock, J.O.: Slide rule: making mobile touch screens acceptable to blind people using multi-touch interaction techniques, pp. 73–80. ACM, New York (2008)
4. Sanchez, J., Maureira, E.: Subway Mobility Assistance Tools for Blind Users. In: Stephanidis, C., Pieper, M. (eds.) ERCIM Ws UI4ALL 2006. LNCS, vol. 4397, pp. 386–404. Springer, Heidelberg (2007)
5. Hutchins, E.L., Hollan, J.D., Norman, D.A.: Direct Manipulation Interfaces, User Centered System Design, 87/124. Lawrence Erlbaum, Mahwah (1986)
6. DAISY consortium, <http://www.daisy.org>
7. PTP1 Users Guide. Shinano Kenshi Co., Ltd. (2009)

Inspired by Sharp Vision: The Full HD Desktop Video Magnifier “Eagle”

Maria Schiewe¹, Thorsten Völkel², and Dirk Kochanek¹

¹ F. H. Papenmeier GmbH & Co. KG, Talweg 2, 58239 Schwerte
{[maria.schiewe](mailto:maria.schiewe@papenmeier.de),[dirk.kochanek](mailto:dirk.kochanek@papenmeier.de)}@papenmeier.de

² PPI AG, Wall 55, 24103 Kiel
t.voelkel@hci-research.de

Abstract. This paper describes Eagle. Eagle marks the beginning of a new video desktop magnifier generation. Equipped with a high-definition camera, it offers a resolution up to 1920×1200 pixels supporting the 16:10 aspect ratio by default. We show how Eagle fulfils well-known requirements of desktop magnifiers combining state-of-the-art technologies. We also emphasise in which way users benefit from this novel development.

Keywords: video magnifier, closed-circuit television, electronic vision enhancement system, high-definition camera, low vision, visual impairment.

1 Introduction

Low vision may affect various visual functions such as visual acuity and contrast sensitivity but also loss of peripheral visual field or vision loss within the central 15 degrees of the field of view. Peripheral vision loss may lead to restrictions in mobility and in scanning the environment [2]. The latter leads to decreased reading ability, in particular with the loss of foveal function reading pace being significantly slower. More errors are made compared to readers with intact central field [4]. Reading ability restrictions can lead to loss of enjoyment in leisure reading. [7] It may as well influence the integration into employment because the ability to work with computers is mostly affected.

Desktop video magnifiers, also referred to as closed-circuit televisions (CCTV), have greatly improved working conditions for people with low vision. The reason is that arbitrary documents can be enlarged continuously. Video magnifiers allow people with low vision to read for pleasure, manage finances or use computer equipped working places. Desktop video magnifiers are applied daily by 30% of all users [9] and are well accepted by the user group [8].

However, strong requirements are imposed on desktop video magnifiers to be usable in professional environments. Resolution is critical because a high resolution allows displaying larger parts of a document. Thus, poor or very small documents benefit as more details can be presented. Furthermore, high contrast is needed to compensate effects of contrast sensitivity. The presentation of enlarged images must be completely flicker-free since peripheral vision is most

sensitive to flickering which greatly affects people with central vision loss. Finally, illumination and colour fidelity is an important feature. Without proper illumination border areas are affected negatively. High colour fidelity ensures uniform colour areas as well as smooth colour gradients.

Additionally, some technical requirements must be fulfilled by state-of-the-art systems. Current computer monitors are produced using an aspect ratio of 16:10 (WUXGA) and offer digital input capabilities (HDMI, DVI). Extended display size offers an improvement as the relation between reading rate and window width is almost linear. This imposes that reading speed will gradually grow with wider windows [3].

Most of currently available desktop video magnifiers do not fulfil all cited requirements, for a thorough overview refer to [1] and [11]. Therefore, we present an overview of our novel Visulex Eagle Pro magnifier. It combines latest technology to overcome the stated drawbacks.

2 Technology

Visulex Eagle Pro is a completely new development with technical improvements that address the needs of the target group. Eagle features a high-definition camera with a 5 megapixels CCD sensor and a 15:1 zoom lense from Carl Zeiss. The integrated light source is a high-power LED generating up to 3000 lux on the document.

Eagle is the first magnifier with full high-definition resolution, i. e. 1920×1200 pixels. Although other systems offer the same monitor resolution as well, the pictures have to be upscaled because of the lower camera resolution. In addition, Eagle can operate on up to 60 full frames per second (progressive scan) using full HD resolution compared to not more than 30 frames per second of other magnifiers currently available on the market. As a consequence, the frame rate is optimally adapted to up-to-date TFT displays. This also enables lag- and blur-free presentation of moving documents and avoids flickering.

Eagle offers a small initial magnification of 1.1 times the original document. Only combined with HD resolution, the perception of as much text as possible is allowed, gaining a broad document overview while still providing a sharp image. This is especially helpful for users with retinitis pigmentosa. Also, many other visually impaired persons benefit from minimum magnification since maximum field of view increases their reading speed [6].

Analogue connections cause image interferences when used with high resolutions. Eagle avoids this effect by providing fully digital signal processing and transmission without any analogue-to-digital conversion and vice versa. Advanced image processing is applied to further enhance the output. In addition to standard techniques like white balancing, Eagles uses adaptive shading correction. This enables quality pseudocolour images even at small magnification without disruptive vignetting caused by lighting or lense errors. When no motion is detected, the image is automatically frozen, thus, providing a picture without any edge flickering.

Motion detection and auto-focus can be turned off when needed. Logically, manual freezing is then possible. Supplementarily, users can customise settings such as zoom, brightness, contrast, and saturation. They can choose from three different colour modes: true colour, black and white, or pseudocolour. The last two modes also offer inversion. Pseudocolour mode additionally provides 8 pre-defined colour schemes and allows 3 user-defined schemes. Eagle is operated by a mouse or trackball attached to the control unit, via on-screen display, or via Windows programme.

It is crucial for an efficient working process to provide a split screen including the magnified document and the operating system's desktop. Eagle is capable of such splitting. Camera and PC output are mixed in Eagle's image processing unit (see Fig. 1). This combined image is directly sent to the monitor, thus, preserving maximum frame rate and resolution without using any of the computer's processing power. All video input and image processing are completely digital allowing the use of any screen magnification software. By way of contrast, other desktop magnifiers use USB, FireWire, or LAN as the transport media for the camera signal. They only provide an overlay of the camera picture. Screen magnifiers are not able to recognise and magnify overlays. Working both with split screen and magnification software causes problems for other systems.



Fig. 1. Scheme of Eagles's split screen function

On purpose, no preliminary user study has been conducted. While user studies seem mandatory for almost anything, there is little sense in doing so for magnifier manufacturers. The reason is the extreme heterogeneity, i. e. it is hard to find two visually impaired people alike. Let alone the clinical picture, results also depend on other factors such as ambient lighting, reading distance and angle, monitor size and resolution, read documents—everything out of the manufacturer's direct scope. Thus, results will likely be without technical significance. Consequently, magnifiers should always provide the highest available resolution, support modern input and output channels, and allow as much user customisation as possible.

3 Summary and Conclusion

In this paper, Eagle, a completely novel developed video desktop magnifier has been presented. Eagle allows wide customization. It can easily be adapted for a great variety of work settings and users. Eagle combines latest technology, full high-definition camera resolution at 60 frames per second progressive scan. It offers superior image quality achieved by advanced digital image processing. As Eagle uses a completely digital signal chain, users are able to operate with a split screen without the necessity of overlays. As a consequence, Eagle supports most settings where standard magnification software is applied.

In a nutshell, Eagle combines the strengths of currently available desktop video magnifiers. Users with low vision are offered maximum flexibility and support for various work settings in professional and private environments.

Acknowledgements. We thank Anne Kochanek for her translation of Eagle's slogan. Visulex Eagle Pro is a product of F. H. Papenmeier GmbH & Co. KG. Two of the authors are employees of this company.

References

1. American Foundation for the Blind: 2008 AccessWorld Guide to Assistive Technology Products, pp. 31–96. AFB Press, New York (2008)
2. Brennan, M., et al.: In their own words: Strategies developed by visually impaired elders to cope with vision loss. *J. Gerontological Social Work* 35(1), 107–129 (2001)
3. Brinker, B., Bruggeman, H.: Visual requirements for reading: The importance of a large field of view in reading with a magnifier. *Journal of Videology* 1, 27–38 (1996)
4. Fletcher, D.C., Schuchard, R.A.: Visual function in patients with choroidal neovascularization resulting from age-related macular degeneration: The importance of looking beyond visual acuity. *Optometry and Vision Sci.* 83(3), 178–189 (2006)
5. Fletcher, D.C., Schuchard, R.A., Watson, G.: Relative locations of macular scotomas near the PRL: Effect on low vision reading. *J. Rehabil. Res. Dev.* 36, 356–364 (1999)
6. Lovie-Kitchin, J.E., Woo, G.C.: Effect of magnification and field of view on reading speed using a CCTV. *Ophthal. Physiol. Opt.* 8(2), 139–145 (1988)
7. Ramrattan, R.S., et al.: Prevalence and causes of visual field loss in the elderly and associations with impairment in daily functioning: The Rotterdam study. *Arch. Ophthalmol.* 119, 1788–1794 (2001)
8. Rinnert, T., Lindner, H., Behrens-Baurnann, W.: Nutzungshäufigkeit vergrößernder Sehhilfen im Wohnbereich von Sehbehinderten. [Home use frequency of low-vision aids by partially sighted]. *Klin Monatsbl Augenheilkd* 215, 305–310 (1999)
9. Rohrschneider, K., et al.: Nutzung und Akzeptanz von vergrößernden Sehhilfen [Satisfaction with low vision aids]. *Klin Monatsbl Augenheilkd* 219, 507–511 (2002)
10. Schoessow, K.A., Fletcher, D.C.: Shifting Desktop Video Magnifier Monitors to Compensate for Central Scotomas. *AER Journal* 2(2), 76–83 (2009)
11. Wolffsohn, J.S., Peterson, R.C.: A review of current knowledge on Electronic Vision Enhancement Systems for the visually impaired. *Ophthal. Physiol. Opt.* 23(1), 35–42 (2003)

Non-sequential Mathematical Notations in the LAMBDA System

Cristian Bernareggi

Università degli Studi di Milano, Biblioteca di Informatica

Abstract. Blind people are used to read mathematical notations through sequential representations (e.g. most Braille codes represent fractions, superscripts and subscripts in a sequential pattern). In mathematics, many notations cannot be represented through a sequential pattern without a loss of suggestivity and expressiveness (e.g. matrices, polynomial division, long division, automata, etc.). This paper introduces a multimodal interaction paradigm which aims to enable blind people to explore and edit mathematical notations which cannot be represented through a sequential pattern. So far, this interaction paradigm has been implemented in the LAMBDA system.

Keywords: Accessibility, mathematics, blind.

1 Introduction

Over the years mathematical expressions have been made accessible to blind people through sequential representations. Although there exist many differences among national Braille systems for mathematics [1], all of them share the same peculiarity to represent expressions through sequential patterns. For example, fractions, which are visually represented through a bi-dimensional layout, are sequentially encoded in Braille through tags which explicitly mark numerator and denominator in a sequential representation. Furthermore, speech renderers for mathematical expressions usually generate speech output of messages made up of the names of symbols in the expression. Nonetheless, certain mathematical notations cannot be sequentially represented without a loss of suggestivity and expressiveness. For example, if a matrix was represented through a sequential pattern (e.g. a comma separated list of elements on the rows and a semicolon separated list of rows), diagonals could not be easily explored and understood thus making it difficult to discover regularities in the matrix. Further, columnar arithmetic operations (e.g. long division), which are taught in the early years of primary schools, could not be sequentially represented because of the vertical and horizontal relations among digits. Traditionally, non-sequential notations, which cannot be translated into a sequential representation without a loss of meaning, have been represented both paper through tactile symbols and Braille characters. These representations, once embossed, cannot be edited any longer and they are constrained to the size of the sheets. This paper, at first, introduces an interaction paradigm designed to enable blind people to read and process mathematical notations which cannot be sequentially represented without a loss of expressiveness and suggestivity.

This interaction paradigm is based on the use of tabular structures, which can be explored through a Braille display and a speech synthesizer in the LAMBDA system [2]. Subsequently, three case studies have been examined: polynomial division, solving inequalities and automata.

2 The Interaction Paradigm

The interaction paradigm to enable blind people to work with certain non-sequential mathematical notations is based on some preliminary remarks.

Tabular structures can be employed to represent many non-sequential mathematical notations. For example, a matrix can be represented in a table, each digit and the operator in a long division can be placed in the cells of a table, graphs can be represented through an adjacency matrix, and so on).

Since a blind person is used to manipulate sequential structures, tabular structures should be encoded in a sequential pattern in the document. So, the blind reader is not forced to change the usual reading style (typically from left to right) whenever a tabular structure is found.

Nonetheless, a sequential pattern may cause the loss of expressiveness and suggestivity [3]. Therefore, an alternative representation together with a set of exploration and editing operations may help also the blind reader to get advantage of expressiveness and suggestivity properties, which are usually conveyed by non-sequential notations. The interaction paradigm proposed in this work is based both on a sequential and on a bi-dimensional representation of tables as well as on a set of primitive editing and exploration operations.

Since many non-sequential notations or procedures based on non-sequential notations (e.g. multiplication of matrices) need be represented in strongly related multiple tables, operations to work with more than one table at a time are indispensable.

2.1 Representation of Tables

The LAMBDA system enables blind users to work with scientific documents. All the parts making up the document (e.g. text and mathematical expressions) are represented through sequential patterns. Therefore, tables are represented in the document through a sequential pattern, consistently with the representation of all the elements in the document. Nonetheless, this sequential representation is not suitable to guarantee expressiveness and suggestivity of bi-dimensional notations. Therefore, an alternative bi-dimensional representation of tables is available in the LAMBDA system. This bi-dimensional representation is available on demand, so that it does not interfere with the usual reading flow. The following examples illustrate the two representations of tables made available in the LAMBDA system.

2.1.1 Example 1

The sequential representation follows:

Row = element_1 column separator element_2 colum separator ... elem_n

Tabular-structure = open tag Row_1 row separator Row_2 row separator ... row separator Row_m close tag

For illustrative purposes, in the following examples it is assumed:
 open tag = { column separator = , row separator = ; close tag = }

$$\begin{vmatrix} 1 & 2 \\ 3 & 8 \end{vmatrix}$$

Fig. 1. Matrix

It is encoded as: {1,2;3,8}

2.1.2 Example 2

In the bi-dimensional representation cell separators are explicitly marked and made not editable, alignment of the content in each cell is preserved also for a Braille display user.

$$\begin{vmatrix} 1 & 2 \\ 3 & 8 \end{vmatrix}$$

Fig. 2. Matrix

is represented as:

Table 1. Table representing a matrix

1	,	2
3	,	8

2.2 Exploration and Editing Operations

This section illustrates basic editing and exploration operations which can be used to work on tables. All the operations have the same behaviour both on the sequential and on the bi-dimensional representation. These operations are widely used in the case studies introduced in the next section.

New table

A new table is created by choosing the number of rows and columns. It is sequentially represented in the document as an empty table. For example, a 2 by 2 matrix is represented as {,;,}. The speech synthesizer will vocalize the names of the markers (e.g. "Open table", "close table", etc.) and the number of rows and columns.

Switch to tabular mode

This operation is applied to a table whenever a representation in a two-dimensional layout is needed. The table is represented in a bi-dimensional layout. Editing operations are available only in the table so that the document cannot be edited synchronously thus avoiding inconsistency. If more than one table are inserted on the same line in the document, all the tables will be available for exploration and editing in tabular mode.

Right/left up/down

The focus is moved to the next/previous cell either on the row or on the column. The content of the cell is read by speech.

Next/previous table

If multiple tables are available in sequential or tabular mode, the focus is moved to the next/previous table. Only one table at a time is displayed. That ensures easier exploration on the Braille line. The position of the focus in the table is the one after the latest operation on the table.

Add/delete row/column

Either a column or a row is added or deleted in the table.

3 Case Studies

3.1 Introduction

This section introduces three case studies which have been examined. Each case study employs the interaction paradigm based on tabular structures to make accessible a non-sequential mathematical notation or procedure based on non-sequential notations, through speech output and Braille display. The choice of each case study was led by a preliminary examination of the scientific school programs in secondary schools and universities.

3.2 Polynomial Division

Polynomial division is a mathematical procedure which employs non-sequential notations. It was chosen as a case study because it is widely taught in secondary schools and university courses. Furthermore, it suggests the applicability of the use of tabular structures to similar cases (e.g. long division, Ruffini method, etc.).

The description follows:

1. The user writes in the document the divisor and the dividend as it is often required by many teachers. Subsequently, to the purpose of applying the procedure of polynomial division, the student creates an empty table. The number of rows and columns equals the grade of the dividend polynomial plus 2 (two more columns are needed to have an empty column between the dividend and the divisor and to write the divisor in the first cell of the rightmost column). However, it is not indispensable for the user to determine the exact number of rows and columns in advance since editing operations to add/delete rows or columns are available;

2. The user switches to tabular mode. Tabular mode is supposed to be far more suitable to apply the polynomial division procedure. Actually, columnar and horizontal relations between polynomial terms must be detected by the student;
3. The user fills in the first row the dividend and the divisor;
4. By using editing and exploration operations, the user performs the polynomial division.

Step 2 and 3 can be swapped according to the user's familiarity with sequential or bi-dimensional exploration.

The following example illustrates the representation in a table of the polynomial division procedure. This representation is displayed in tabular mode.

3.2.1 Example 3

Remarks

The user might have created two empty tables on the same line at the beginning of the procedure. They would be both available in tabular mode. So the user might have exploited a third dimension to add comments related to the main table. By using next/previous table operations each comment could be easily related to the corresponding cell.

Here is an example related to the polynomial division previously illustrated.

Table 2. Table representing the polynomial division procedure

x^2	0	-1		$x+1$
x^2	0	0		x
0	0	-1		

Table 3. Labels for the polynomial division procedure

dividend	dividend	dividend		divisor
				quotient
		remainder		

3.3 Solving Inequalities

This section illustrates a method to solve inequalities by using a table. This method was chosen as a case study because it is widely employed in secondary schools.

A description follows:

1. The user simplifies the inequality and factorizes it. Once factorized, the sign rule is applied;
2. The user creates an empty table. The number of rows and columns equals two times the number of factors plus 1 (e.g. it equals the number of intervals determined by the roots of the factorized expression). Nonetheless, the user is not required to create the table with the right number of columns and rows. Editing operations to add/delete rows/columns can be employed whenever needed to adapt the number of rows or columns in the table;

3. The user switches to tabular mode;
4. The user writes the limits of the intervals in the even cells on the first row and the signs in the odd cells from the second row on;
5. The user applies the sign rule and writes the results in the last row.

An example follows

3.3.1 Example 4

$$x^2-x-6>0 \quad \rightarrow \quad (x-3)(x+2)>0 \quad \rightarrow \quad x>3 \text{ or } x>-2 \quad (1)$$

Table 4. Table for inequality solution

	-2		3	
-	0	-	0	+
-	0	+	0	+
+	0	-	0	+

$$x<-2 \text{ or } x>3 \quad (2)$$

3.4 Automata

This section illustrates how to represent and explore automata through tabular structures. Automata were chosen as a case study because they deserve a very important role in computer science courses and their representation can suggest the applicability of the method to other notations based on directed graphs. A more comprehensive work on automata representation is available in [4],[5]. Automata are represented through an adjacency matrix described as follows:

- The labels of automaton states are the headers of rows and columns;
- A full stop (.) is placed before the label of the start state and a full stop (.) is placed after the label of each accepting state;
- Symbols from the alphabet for the language accepted by the automaton are placed in the matrix cells to define the transition function. A symbol *s* placed in the position (i,j) is interpreted as a transition from the state labeled as in the header of the i-th row to the state labeled as in the header of the j-th row through symbol *s*. If a transition is achieved by more than one symbol, a comma separated list of symbols is placed in the corresponding cell;
- If no transition exists between two states, the corresponding matrix cell is left empty.

The user can employ the operations in tabular mode to explore and edit the automaton.

4 Conclusions

This paper has illustrated an interaction paradigm to explore and edit non-sequential mathematical notation in the LAMBDA system. The examined case studies suggest

further applicability of this paradigm. Up to now a formal assessment has not been undertaken, nonetheless the introduced strategies have been employed by three secondary school blind students and 2 university blind students in computer science.

References

1. Gopal, G., Stoger, B., Palmer, B., Guo, H.: UMA: a System for Universal Mathematics Accessibility. In: ASSETS 2004, ACM Press, New York (2004)
2. Schweikhardt, W., Bernareggi, C., Jessel, N., Encelle, B., Gut, M.: LAMBDA: a European System to Access Mathematics with Braille and Audio Synthesis. In: Miesenberger, K., Klaus, J., Zagler, W.L., Karshmer, A.I. (eds.) ICCHP 2006. LNCS, vol. 4061, pp. 1223–1230. Springer, Heidelberg (2006)
3. Iverson, E.: Notation as a Tool of Thought. In: SIGAPL APL Quote Quad 35, ACM Press, New York (2007)
4. Stallmann, M.F., Balik, S.P., Rodman, R.D., Bahram, S., Grace, M.C., High, S.D.: Proof-Checker: an Accessible Environment for Automata Theory Correctness Proofs. In: ITiCSE 2007: 12th Annual Conference on Innovation & Technology in Computer Science Education, Inclusive Education in Computer Science, ACM Press, Dundee (2007)
5. Bernareggi, C., Dalto, G.C.: Making Automata Accessible to Visually Impaired People. In: WEIMS conference, Fukuoka (2009)

MathML to ASCII-Braille and Hierarchical Tree Converter

Silvia Fajardo-Flores, Maria Andrade-Arechiga,
Alfonso Flores-Barriga, and Juan Lazaro-Flores

University of Colima, Mexico
{medusa,mandrad,alfonso_flores,juan_lazaro}@ucol.mx

Abstract. Mathematics is a complex study subject for most people. People with disabilities, specifically those who are blind, face additional challenges: lack of study material and difficulty to access digital math notation. Significant efforts have been made in order to improve the quality of access to materials; however a complete solution has not been provided; in addition to this, most software applications support European and North American mathematical Braille codes, but very few of them cater for Spaniard and Latin American users with blindness. The present paper describes a prototype that converts equations in MathML into two alternative formats for representing them: an ASCII-Braille file ready for embossing, and a navigable tree structure conveying the hierarchy of equations that can be read with a screen-reader. The prototype uses the Unified Mathematical Code. The current version of the prototype presented here is functional and can be used for producing material for people with blindness; however, it still requires testing with final users.

1 Introduction

1.1 Background

People with blindness face serious difficulties studying mathematics and science related areas because of the limited access to materials both in printed and digital format, the difficulty reading and writing complex mathematical notation, and the generally poor interaction with their sighted teachers due to differences accessing the same information.

Archambault and Stöger mention some challenges faced by people with blindness studying Mathematics [1]: a) The biggest barrier is not understanding Mathematics, but accessing content, which should be available in Braille or speech. On this regard, Maddox states: "Mathematics is not easy to convey using Braille or screen readers and although partial solutions exist, they still have limitations" [2]; b) Increasing number of students are attending mainstream schools, therefore it is important that tools are accessible to teachers without Braille knowledge. Miesenberger regards as a challenge to present, communicate and

cooperate with mathematical information [3]; c) Bidimensional nature of Mathematical expressions, while Braille and speech are linear; therefore formulas need to be linearised.

According to Miesenberger, there is yet one more challenge: allowing blind people to do mathematical calculations [3].

Braille Math. Braille is still the most pervasive format for people with blindness. Literary Braille is standard across languages using the Roman alphabet; however, codes for Math vary amongst different countries. Examples of Braille math codes are:

- Nemeth: Used mainly in the US, but it is also accepted in other countries such as Canada, India, and Malaysia.
- RNIB: Used in the UK
- Marburg: Used in Germany, Austria and Poland
- CMU (Unified Mathematical Code): The math standard Spanish Braille for Latin America and Spain

Nemeth states that the pervasiveness and effectiveness of Braille is still not comparable to the use of modern technologies [4].

1.2 Software Solutions

Blind people are able to use computers with the aid of a screen reader, special software that provides speech output. Screen readers are very effective reading literary Braille; however, they present limitations reading Mathematical expressions both on desktop applications and the Web.

Different software applications have been developed, and methods and strategies have been proposed to alleviate this situation. To date, the techniques developed to aid blind people in the study of Mathematics may fall within the categories of: tactile (including Braille), audio aids, tonal representation, haptic devices, or an integration of at least two of these [5].

There are several applications aiming to support the translation of Math to and from Braille, some of which include speech output. MAVIS and LaBraDoor translate equations to Braille; from those, MAVIS provides audio output through the screen reader and uses a hierarchical structure to sub-group equations into blocks. Perhaps the most complete system for working with Braille math is LAMBDA, which converts equations into different Braille formats using 8 dots and includes audio output; however, it does not build a hierarchical structure of the equations.

The W3C Math Working Group launched MathML, an XML vocabulary for marking mathematics, with two sub-languages: presentation MathML and content MathML. Bernareggi and Archambault state that the MathML recommendation should be adopted in order to make Web documents fully accessible to both blind and sighted people, and to be read with screen readers or Braille devices [6].

Tools such as the Microsoft Word Equation Editor allow users to produce complex Mathematical formulas for printed documents; however, direct conversion to Braille is not as straightforward [7].

The applications mentioned before are available mainly in mathematical codes used in the US and some European countries, but do not include conversion to CMU. On this regard, Alonso et al. developed a translator from Spanish mathematical Braille to MathML, using ASCII-Braille as source code for Math expressions to produce a visual representation of equations, opening an opportunity for communication between blind and sighted people [8].

1.3 Methods and Strategies

Karshmer, Pontelli and Gupta mention that since the majority of Math teachers are sighted and do not know Braille, it is necessary an application that allows interaction between sighted teachers and blind students, making it easier for both to read and write in the format they are used to [9].

On the same line, Archambault and Stöger suggest as a strategy to design tools to support collaborative work between blind and sighted in a mainstream learning environment, which requires synchronization of two representations using two different modalities according to each user (visually impaired and sighted) natural view [1]. Bernareggi and Archambault report on a prototype named MaWEn (Mathematical Working Environment), a bi-modal software that enables users to read and write equations, collapse and expand its terms, and navigate through them in a synchronized view; the MaWEn prototype has been tested in laboratory conditions [6]. This software uses a Universal Maths Conversion Library (UMCL) in order to support several math codes; however, the literature is not clear regarding inclusion of the CMU in the library.

Fitzpatrick proposes methods for depicting Mathematical formulas in a non-visual manner; he suggests that it is possible to disambiguate a formula by conveying its structure through the use of pauses and alterations in the speaking rate [11]. He suggests the combination of this audio technique with touch-screen technology in order to produce a multi-modal presentation of complex data.

2 Converter Prototype

2.1 Overview

The current proposal consists of a prototype to convert MathML algebraic expressions to:

- A linear representation in an ASCII-Braille file ready for embossing, using CMU.
- A navigable hierarchical representation of equations that allows collapsing and expanding branches [9,10].

Figure 1 illustrates the functionality of the converter.

As a first prototype, the development of the converter was focused on the output; that is to say, in producing the ASCII-Braille file and the hierarchical

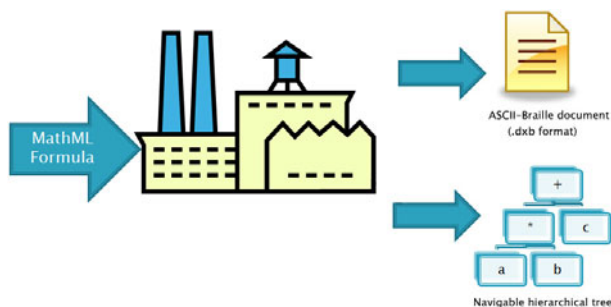


Fig. 1. Converter Input/Output data

tree. The interface at the moment is not user-oriented, since it requires MathML knowledge in order to provide an input. In a second phase, a user-friendly interface will be designed to cater for non-technical users.

This prototype converts basic algebra equations for the following operations:

Operation	Symbol
Addition	+
Subtraction	-
Multiplication	*
Division	/
Root	$\sqrt{\quad}$
Exponent	\wedge

2.2 Implementation of the Prototype

The prototype was developed in Visual Basic as a desktop version that requires an equation in Content MathML as input; Content MathML was preferred over Presentation MathML, since it conveys the meaning of the equation instead of its layout, resulting in an unambiguous representation that can be easily translated into CMU. The prototype works as follows (see table 1 for an example): Following the structure of the markup code, the converter transforms the MathML labels into an XML format, then parses the XML structure in order to generate a prefix tree of the equation using the TreeView control, and then converts the prefix tree into infix in order to represent the equation in a reading format that is familiar for people. The structure of the hierarchical infix tree is displayed in the interface, and can be read by screen-reading software. Users are able to expand or collapse branches using the + and keys, and to navigate the structure using the arrow keys or the mouse. The infix structure shown in Table 1 should be the final layout to be displayed in the interface; however, it requires a script for the screen-reader software in order to read expressions adequately. As a temporary fix alternative to the script, the operators in the tree are converted to actual words in order to "speak" the language of the user.

In order to produce the ASCII-Braille output for embossing, the infix tree is traversed while each character is searched in a table for its ASCII-Braille equivalence, according to the CMU. The characters in ASCII-Braille are written into

Table 1. Example of functionality of the prototype

<p>Original equation</p> $a + b \times 2^{\frac{1}{3}}$	
<p>1) Content MathML input</p> <pre><math xmlns="http://www.w3.org/1998/Math/MathML"> <apply> <plus> <ci>a</ci> <apply> <times> <apply> <power> <cn type="integer">2</cn> <cn type="rational">1<sep>3</cn> </apply> <ci>b</ci> </apply> </times> </apply> </plus> </apply> </math></pre>	
<p>2) Process</p> <p>Tree hierarchy (prefix)</p>	
<p>3) Output</p> <p>Infix structure</p>	
<p>Temporary equation wording in Spanish for screen-reading software</p>	
<p>ASCII-Braille: a655#b*#1c98b9</p> <p>Embossed text:</p>	

a text file that can be saved as a Duxbury file (.dxb) and embossed in a Braille device. The ASCII-Braille file was tested using Duxbury 10.7 and Winbraille 4.0, and embossed in an Index-Braille Basic-D.

2.3 Test Plan

The test of the prototype has not been completed. The prototype will be tested in three aspects: 1) Compliance of the ASCII-Braille file with CMU rules; 2) User testing for conveyance of the structure of equations using the hierarchical tree; and 3) Usability of the tree structure.

PARTICIPANTS

For (2) and (3), the participants for this test will be people with blindness with basic knowledge of Algebra, and two passive observers of the conduction of the tests who will take notes on specific issues that may be found.

MATERIALS

Several equations in MathML will be prepared for the tests, from the simple to the complex, in order to generate the output tree for the user. Participants will be using a PC with the JAWS 10 screen-reader.

PROCEDURE

The procedure for each test is planned as follows:

1. Compliance of the ASCII-Braille file with CMU rules. This is a black box test and does not require user participation; the ASCII-Braille output generated from the MathML equations and saved in a DXB file will be embossed and then evaluated for compliance with the CMU. A preliminary test has been done with operations such as polynomials and equations with expressions in numerator/denominator, and no compliance issues with CMU were found, but it is required a thorough evaluation with more complex equations.

2. User testing for conveyance of the structure of equations using the hierarchical tree. It is necessary to find out if the user with blindness understands the structure of equations from navigating the tree. The user will be asked to navigate through the output tree of each equation and then asked to write it in Braille using the tool they are familiar with (i.e. slate and stylus, Perkins machine); they will be able to check on the tree while writing the equation. The resulting written expression will be reviewed according to the CMU, and compared with the original structure.

3. Usability of the tree structure. While testing for conveyance of the structure, it is necessary to find out how easy it is for the final user to navigate through it, and the issues that he or she may find, in order to improve the user experience. For this purpose, we will be using the Thinking aloud protocol, that requires giving the user specific tasks, and asking him or her to verbally express the feelings and intentions while doing them. Each user will be recorded in video, and notes will be taken during and after the conduction of the test.

The results of the tests will serve as feedback for an improved version of the prototype.

3 Conclusions

The prototype presented in this document is a functional converter that requires MathML code as input, and produces an ASCII-Braille document and a hierarchical tree for use with a screen reader as output. The output ASCII-Braille is compliant with the CMU and is ready for embossing. Once a non-technical user interface is designed, users with no knowledge of MathML will be able to produce material; however, at the moment the prototype is limited to algebraic equations. The hierarchy layout requires testing with blind users in order to find out if the proposed structure conveys the equation and if the navigation suits them; it is expected that users will provide useful information for an improved version of the prototype.

References

1. Archambault, D., Stöger, B.: Mathematical working environments for the blind: What is needed now? In: Seventh IEEE International Conference on Advanced Learning Technologies, ICALT 2007 (2007)
2. Maddox, S.: Mathematical equations in Braille. *MSOR Connections* 7(2) (May-July 2007)
3. Miesenberger, K.: Doing mathematics. @Science Thematic Network (2008), <http://www.ascience-thematic.net/en/conferences/paris/Miesenberger>
4. Nemeth, A.: Teaching Mathematics as a blind person. The Text Presentation to the Mathematical Association of America, Orlando, Florida, January 10 (1996)
5. Karshmer, A., Bledsoe, C.: Access to Mathematics by blind students. In: Proceedings of the 8th International Conference on Computers Helping People with Special Needs, Linz, Austria (2002)
6. Bernareggi, C., Archambault, D.: Mathematics on the Web: emerging opportunities for visually impaired people. In: Proceedings of the 2007 international cross-disciplinary conference on Web accessibility (W4A), Banff, Canada, pp. 108–111 (2007)
7. Mani, M., Goswami, S., Dutta, S., Datta, S.: Mathematics Braille. *The Educator* 18(2) (January 2006)
8. Alonso, F., Fiertes, J.L., González, A.L., Martínez, L.A.: SBT: A Translator from Spanish Mathematical Braille to MathML. In: Miesenberger, K., Klaus, J., Zangler, W.L., Karshmer, A.I. (eds.) ICCHP 2006. LNCS, vol. 4061, pp. 1207–1214. Springer, Heidelberg (2006)
9. Karshmer, A., Pontelli, E., Gupta, G.: Helping visually impaired students in the study of Mathematics. In: 29th ASEE/IEEE Frontiers in Education Conference, San Juan, Puerto Rico, November 10-13 (1999)
10. Archambault, D., Stöger, B., Batusić, M., Fahrenguber, C., Miesenberger, K.: A software model to support collaborative mathematical work between Braille and sighted users. In: Ninth International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS 2007), Tempe, Arizona, October 15-17 (2007)
11. Fitzpatrick, D.: Teaching science subjects to blind students. In: Seventh IEEE International Conference on Advanced Learning Technologies, ICALT 2007 (2007)

Tactile Graphic Tool for Portable Digital Pad

Thamburaj Robinson^{1,2} and Atulya K. Nagar²

¹ Department of Mathematics, Madras Christian College
Chennai 600 059, India

robin.mcc@gmail.com, thambur@hope.ac.uk

² Department of Computer Science, Liverpool Hope University
Liverpool L16 9JD, United Kingdom
nagara@hope.ac.uk

Abstract. Hand exploratory movements are an important aspect in the pedagogical exercise of teaching and learning diagrammatic representations among vision impaired students. The Tactile Graphic Tool is one such device allowing hand exploratory movement in making tactile diagrams of graphical and geometric constructions. In this paper the same device is modified to be used along with a portable digital pad making the tactile picture drawn accessible as digital picture in computer through its interface. Illustrations of tactile geometrical construction in tactile and digital format are provided.

Keywords: Tactile diagrams, visual impairments, mathematics, accessibility.

1 Introduction

The graphical information contents such as graphs, statistical representations, geometry constructions and the technical, scientific on-line documents that involve mathematical expressions cause a great challenge to the people with visual impairment. The challenge is much greater for them while reproducing or drawing such representations. With the advent of computer assisted technology, there has been an array of tools with multi-modal approach. TDraw, one of the earliest developed by Kurze [4] used a Thermostift digitising tablet and heat sensitive swell paper as a method of input. Speech recognition was used to simulate keyboard commands. The drawing was constrained to polygons and lines or sequences. Kamel et al [1] proposed a drawing program called the Integrated Communication 2 Draw in which a keyboard was used to navigate the grid, select points and draw shapes. This grid-based model assisted the users in exploring graphics by giving a mental model for the position of objects. Sugimoto et al [10] used an acrylic pen input guide fitted on a tablet with sound localization technique and examined the possibility of drawing formal basic line figures. A tactile display system MIMIZU equipped with a pointing device was developed by Kobayashi et al [3]. With a stylus pen device, a computer equipped with an LCD tablet, and a matrix arrangement of pins as tactile display unit the user could draw and erase tactile images. It was found to be useful not only in tactile drawing, but also as an interactive multimedia displaying method. More recently, the

AutOMathic Blocks system designed by Karshmer [2] combines Braille labelled manipulative blocks assisting the learning of arithmetic and algebra.

On the other hand, in the area of teaching and learning of shapes, the experiment conducted by Wexler [11, 12] with visually impaired students demonstrated that the concepts of shape are the outcome of tactual and kinesthetic experiences. These shapes can be grasped by both hands and recognized by active stimulation or by being held in one hand and manipulated by the other. According to Mani [5], the key factors for effective learning of Mathematics include usage of appropriate teaching aids and provision of simulating experiences, creation of situational approaches etc. Miller [6] observed that perception depends jointly on touch and movement information from active exploration by fingers and palms of one hand and passive tactile information from the other hand. While learning graphs, geometry constructions and statistical representations, the progressive stages need to be understood with exploratory practical experiences. Preparation of tactile graphics using swell paper, Thermoform sheet, a semi-transparent plastic sheet (also known as german sheet) [9] are commonly used means for making raised line drawings. While teaching the intermediate stages of geometry constructions, the usage of 'progressive step method' [8] on swell papers was found to be complementary in drawing tactile constructions using the 'Tactile Graphic Tool' (TGT) experimented by Robinson [7]. It was found that TGT provided the avenue for exploratory hand movements in making of tactile diagrams. TGT is modified further as a tool mountable on a portable digital pad. In this paper, the modification which facilitates the diagram drawn to be stored simultaneously as a digital file is discussed.

2 Tactile Graphic Tool

It is a set of white coloured templates, each 2 mm thickness with, Braille and black tactile markings. There are pin-holes with tactile indications and v-notches at specific positions. TGT consists of an A4 size rectangular tactile annular frame, a rectangular perforated tactile grid template (set to fit inside the annular frame), a tactile square protractor of size 10cm, a tactile scale of 29cm long, a circular disc of 12 cm diameter and a tactile compass scale of 10cm long. Both annular frame and grid template have Braille and tactile markings on both the sides.

TGT makes use of the principle that when a stylus or sharp pen is used on a semi-transparent plastic sheet (german sheet) kept on rubber mat, raised lines are produced. Flat board pins kept at desired positions of the templates or the frame act as position indicators and a tactile line joining two positions could be drawn by the tactile scale. A line or a curve thus drawn protrudes enough for the tactile sense. The annular frame or template can be fixed on a rubber mat (of 3 mm thickness) and cork sheet sandwiching the german sheet.

The subjects of the case studies conducted [7] made tactile quadrilaterals, polygons and geometry constructions using TGT. The annular frame and circular disc templates are used while making statistical diagrams such as pie-chart, histogram while rectangular grid is used for plotting of curves and graphs.

3 Modified Tactile Graphic Tool

The digital pad (DigiMemo A402) is a device with storage capability that captures the scripts written or diagrams drawn on its surface using ink. The content written or drawn thus can be transferred, organised or edited when connected to a computer. TGT is modified (295 mm X 200 mm) with its annular frame to fit on the A4 size area of the digital pad. The cork sheet of 5 mm thickness is used as there is a limitation for the pad to sense the inking pen with 14 mm proximity. Apart from flat headed pins, folded pins are used to mark coordinate positions of the annular frame. The digital pad can be used a stand-alone device or as a tablet device (fig.1) which instantly synchronises the drawing done on it. The storing capacity on number of pages varies depending on the memory storage device. The digitally stored picture of the pad is then imported through pre-installed manager software in the PC as an e-book file. The picture thus saved can further be edited, resized or saved as image file type. Figure 2.a shows the geometrical construction of bisecting an angle of 130 degrees as an e-book digital picture. The rotated and edited version of the same into a MS-Word document is shown in fig.2.b.



Fig. 1. Modified Tactile Graphic Tool mounted on a Digital pad

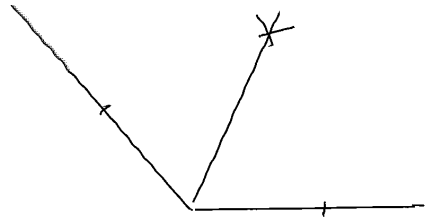
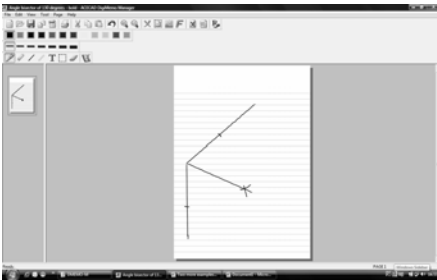


Fig. 2. Construction of angle bisector (a) screen shot as an e-book file (b) imported file into MS-Word document

4 Conclusion

This paper discussed the coupling of a classroom geometry tool with a portable digital pad for making tactile and digital pictures simultaneously. Using of segregated regions or quadrants of the digital pad, manipulative pins with speech feedback for precise graph plotting and curve tracing is under investigation. This portable system is cost effective and promises to be feasible for an inclusive education as well.

Acknowledgments. The authors would like to thank the referees for useful comments. They would also like to acknowledge the support rendered by the Leverhulme Trust of the UK.

References

1. Kamel, H.M., Landay, J.A.: A Study of Blind Drawing Practice - Creating Graphical Information without the Visual Channel. In: 4th Annual ACM Conference on Assistive Technologies (ASSETS 2000), pp. 34–41 (2000)
2. Karshmer, A.I.: Using AutoMathic Blocks to Teach Math Braille to Children (April 6, 2010), http://river-valley.tv/media/conferences/math_for_blind_2008/0103-Karshmer/
3. Kobayashi, M., Watanabe, T.: Communication System for the Blind Using Tactile Displays and Ultrasonic Pens – MIMIZU. In: Miesenberger, K., Klaus, J., Zagler, W.L., Burger, D. (eds.) ICCHP 2004. LNCS, vol. 3118, pp. 731–738. Springer, Heidelberg (2004)
4. Kurze, M.: TDraw - A Computer-based Tactile Drawing Tool for Blind People. In: 2nd Annual ACM Conference on Assistive Technologies (ASSETS 1996), pp. 131–138 (1996)
5. Mani, M.N.G.: Techniques of Teaching Blind Children. Sterling Publishers, New Delhi (1992)
6. Miller, S.: Understanding of Representing Space, Theory and Evidence from Studies with Blind and Sighted Children. Clarendon Press, Oxford (1994)
7. Robinson, T.: A Drawing Device of Tactile Pictures for the Visually Challenged Students. *Jpn. J. Disability Sciences* 31, 171–178 (2007)
8. Robinson, T.: The Progressive Step Method for Teaching Geometry Constructions to Students with Visual Impairments. *Jpn. J. Spec. Educ.* 44(6), 543–550 (2007)
9. Sheppard, L., Aldrich, F.K.: Tactile Graphics - A Beginner's Guide to Graphics for Visually Impaired Children. *Primary Science Review* 65, 29–30 (2000)
10. Sugimoto, M., Itoh, K., Shimizu, M.: Simple Pattern Creation Using a Pen Input Guide and Sound Localization for the Visually Impaired. In: Miesenberger, K., Klaus, J., Zagler, W.L., Burger, D. (eds.) ICCHP 2004. LNCS, vol. 3118, pp. 745–752. Springer, Heidelberg (2004)
11. Wexler, A.: Experimental Science for the Blind – An Instruction Manual. RNIB publication, Pergamon press (1961)
12. Wexler, A.: Shape Recognition and Drawing by the Blind. *New Beacon* 49(581), 228–233 (1965) and 49(582), pp. 254–258

Spoken Mathematics Using Prosody, Earcons and Spearcons

Enda Bates and Dónal Fitzpatrick

Dublin City University, Dublin 9, Ireland
dfitzpat@computing.dcu.ie

Abstract. Printed notation provides a highly succinct and unambiguous description of the structure of mathematical formulae in a manner which is difficult to replicate for the visually impaired. A number of different approaches to the verbal presentation of mathematical material have been explored, however, the fundamental differences between the two modalities of vision and audition are often ignored. This use of additional lexical cues, spatial audio or complex hierarchies of non-speech sounds to represent the structure and scope of equations may be cognitively demanding to process, and this can detract from the perception of the mathematical content. In this paper, a new methodology is proposed which uses the prosodic component found in spoken language, in conjunction with a limited set of spatialized earcons and spearcons, to disambiguate the structure of mathematical formulae. This system can potentially represent this information in an intuitive and unambiguous manner which takes advantage of the specific strengths and capabilities of audition.

Keywords: Math, auditory interfaces, visual impairment, earcons, spearcons.

1 Introduction

Printed mathematical equations provide a significant amount of information in a highly succinct manner. The visual representation immediately and unambiguously indicates structural information such as the presence and scope of a fraction or the terms contained within a square root operator. Developing a similarly efficient and unambiguous representation for the visually impaired is a significant challenge. Tactile representations such as Braille represent one entirely valid approach to this issue, however, this paper will concentrate on the primary alternative to Braille, speech synthesis.

2 Visual and Audible Representations of Mathematical Material

Printed notation provides a persistent visual cue which is effectively non-temporal, however, audible representations are necessarily fleeting due to the inherently temporal nature of sound. Sighted users can therefore utilize printed material as a form of external memory and do not need to memorize the structure and layout of an equation. This conclusion is supported by the results of a series of cognitive experiments

examining equation reading in sighted users, conducted by Gillan *et al* [1], which found that sighted subjects process operators and numbers more intensively than parentheses. This is perhaps unsurprising as the spatial structure of the equation (which is implied through the use of parenthesis and other graphical symbols and delimiters) is unambiguous and persistent when presented visually. This suggests that working with mathematical material in a non-visual medium will result in an inevitable increase in cognitive load as this structural information must now be held in memory. This strongly implies that any method of presenting the spatial structure of an equation via audition must be as easy as possible to cognitively process.

2.1 Active Browsing of Mathematical Material

Printed mathematics provides a strong spatial definition of the mathematical content and the user provides the temporal dimension by actively directing their attention back and forth through the material [1]. A simple audio representation works in reverse, as the serial stream of audio is highly temporal and the user must now infer and remember the spatial structure from the presented material. In effect, a printed symbol exists outside of time, and a sighted user can concentrate on this symbol whenever they wish. Audition on the other hand is highly temporal and omni-directional, and cannot be consciously directed in the same way as vision. Clearly this functionality is critical if a visually impaired user is to have effective control over the temporal representation of the mathematical material, and indeed the majority of the available software for the presentation of spoken mathematics incorporate some form of active browsing which allows visually impaired users to actively browse and segment mathematical equations [1,2,3,4].

2.2 Spatial Representations

Various projects have attempted to directly map the visible spatial structure of printed notation to an equivalent audible structure. It has been suggested that this will reduce the mental effort required by the user to process and solve the equation [5], however, this may not in fact be the case. The results of a number of experiments have found that the minimum detectable changes in position for visual and auditory stimuli are approximately 1' for visual targets and 2°-4° for auditory targets [6]. The spatial resolution of vision is clearly therefore much more accurate than audition and this suggests that accurately replicating the spatial layout of a printed equation with spatial audio will be difficult to achieve.

While dynamic sonic trajectories can be used to create sonic shapes, this appears to be a cognitive process, rather than a perceptual one [7]. In effect, the shape is cognitively inferred from our perception of the moving sound and is not perceived directly. Therefore it cannot be assumed that mirroring the spatial layout of a written equation with spatial audio will reduce the cognitive load on the listener. In fact, it has been found that the additional mental processing required to determine the spatial trajectory detracts from the processing of the mathematical content [8]. The vertical layout of mathematical equations would appear to be well matched to a spatial audio presentation of an equation, however, auditory localization is particularly inaccurate in the vertical dimension and extremely difficult to synthesize, particularly with binaural

spatialization techniques based on the head-related-transfer function, or HRTF [9, 10]. Poor results have also been found when HRTF techniques are replaced with real loudspeakers positioned vertically in front of the listener [10].

These issues suggest that a linear spatial mapping from vision to audition is not particularly useful due to the fundamental differences between the two senses, however, spatial audio may be beneficial in other ways. Numerous studies have shown that sounds produced from different spatial locations are easier to distinguish, which suggests that if additional sounds are added to the main speech signal, these should be produced from different spatial locations as in the system developed by Goose *et al* [8]. The externalization effect which generally occurs when binaural material is presented via headphones has also been found to be much less fatiguing than standard stereo, and this may also be of some benefit [9].

3 Existing Approaches to Spoken Mathematics

3.1 Lexical Cues

The use of additional spoken cues to indicate the structure of equations and formulae is reminiscent of the linear representation of mathematical material in computer programming languages (which also lack the two dimensional spatial component of traditional printed mathematics). Nemeth MathSpeak is one system which adopts this approach and concentrates on removing the ambiguities which arise in spoken mathematics [11]. MathSpeak uses additional lexical cues such as “*begin fraction*” and “*end-fraction*” (abbreviated to “*B-frac*” and “*E-frac*”) to clearly mark the beginning and end of the fraction. Although this solution entirely removes the ambiguity from spoken mathematics, it becomes confusing as the equations become more complex. For example, the formula for quadratic equations would be described as, “*B-frac minus B plus or minus B-rad B sup 2 base minus 4 AC E-rad over 2 A E-frac*”. In addition, there is evidence to suggest that if the end of an utterance is marked with a verbal message, than this message may detract from the listener’s memory of the utterance, a process which is referred to as the suffix effect [12]. Although Nemeth’s approach largely eliminates structural ambiguity in the presentation, the cognitive effort required to remember and parse the equation is quite significant for all but the simplest material.

3.2 Earcons, Auditory Icons and Spearcons

In contrast to a purely lexical approach, a number of systems have been developed which use additional, non-speech sounds to indicate various aspects of the mathematical material, such as the structure of an equation. Non-speech sounds have long been used to improve access to graphical user interfaces and are often classified as earcons, auditory icons or spearcons. Earcons are generally constructed from abstract sounds such as the error beep in the Windows OS and may consist of melodic or rhythmic musical material. Although the exact meaning of earcons must be learned by the user, the abstract nature of the sound means that earcons are relatively easy to synthesize and modify. In contrast, auditory icons use sampled sounds which are intended to resemble or refer to the object they represent. This means that auditory

icons are more difficult to manipulate than earcons but their meaning may be easier for the listener to recognize. Spearcons are a more recent development and are created by time compressing a spoken phrase so that the resulting short sound is not entirely comprehensible as speech [13].

Using auditory icons to represent structural mathematical constructs like parenthesis or fractions is inherently difficult as there is no obvious relationship between this abstract mathematical syntax and real-world sounds. As it is relatively straightforward to create a hierarchical range of earcons, these may be more appropriate to represent nested parenthesis or other structural notation. This approach was adopted by the MathTalk system which uses musical earcons to indicate structural delimiters (such as a pattern of 3 short ascending and descending notes to indicate the beginning and end of an expression respectively) and also provide an abstract overview of the entire equation [14]. The significant drawback to this approach is that a complex and highly specific musical grammar must be learned prior to using the system. In addition, remembering and decoding musical patterns may be quite difficult for non-musicians and again, the additional cognitive effort required to decode each pattern could detract from the processing of the mathematical content.

Spearcons lie somewhere between recognizable speech and abstract sounds such as earcons. A number of studies have found that spearcons were easier to learn and resulted in an increase in performance in interactive interface tasks such as menu browsing [13]. The major advantage of spearcons is that they can potentially function as either a descriptive lexical phrase or an abstract sound depending on its familiarity to the listener or its function as a structural delimiter.

3.3 Prosody

Prosody, i.e. the patterns of stress and intonation in a language, can also be used to indicate the structure of an equation through the manipulation of parameters such as pause duration, speaking rate and pitch. Raman's AsTeR system used prosody in conjunction with additional lexical cues such as "*fraction*" and "*end-fraction*", and "*quantity*" and "*end-quantity*" as parenthesis to indicate sub-expressions [2]. Different levels of exponents and nested expressions were indicated using persistent speech cues, such as a speaking voice with raised pitch to indicate an exponent. Other systems have attempted to use prosodic schemes more reminiscent of normal speech. Stevens analyzed recordings of spoken algebraic equations and used the pitch, timing and amplitude characteristics of these recordings as the basis for an audio interface designed specifically for the presentation of mathematics [4]. Fitzpatrick argues that the prosodic model used by Stevens conflicts with current research in prosodic and intonational phonology, and instead proposed an alternative model which relates the structure found in mathematical expressions with the inherent composition and structure of the English language [15]. In this model, the nesting of clauses within sentence structures is equated to the nesting of sub-expressions within a larger equation whose structure is similarly indicated through the insertion of pauses and alterations to the speaking rate. Additional lexical cues are added to indicate certain specific terms such as fractions and superscripts, in a similar fashion to Nemeth MathSpeak [11].

Experiments with these systems have found that prosody can be used to determine the structure of an equation and, significantly, required less effort from the user than

lexical cues [4]. Natural prosody would therefore appear to be highly suitable for the presentation of structural information as the subtle prosodic inflections of natural speech are intuitively understood and relatively undemanding in terms of cognitive effort. The significant drawback to this approach is that the structure of complex, nested equations is difficult to determine based on prosody alone, suggesting that some form of additional delimiter is required [4, 15].

4 Prosody and Audio Punctuation, Earcons and Spearcons

The prosodic model developed by Fitzpatrick [15] based on the natural structures found in spoken English appears to be the most cognitively efficient means of conveying the structure of mathematical content via audition. However, in this new model, the prosodic grouping is reinforced with a limited number of additional sounds which function as a form of audio punctuation. These include simple, beep sounds which are used as audio parenthesis to delimit sub-expressions and reinforce the segmentation implied by the prosodic model. The additional lexical cues used by Fitzpatrick to indicate fractions and exponents are also replaced with equivalent spearcons whose shorter duration should reduce the negative influence of the suffix effect. It is beyond the scope of this paper to describe how the prosodic model caters for each individual construct found in mathematical material, however, the following sections will describe how some of the most common terms and expressions will be represented.

The prosody model developed by Fitzpatrick assumes that a mathematical expression is a structure containing an arbitrary number of nested sub-expressions, which can be resolved into either a sentence structure containing various clauses, or a paragraph of sentences [15]. The degree of nesting in the equation is therefore highly important, and once this factor has been determined, the appropriate pauses and changes in speech rate can be determined and added to the basic verbalization of the mathematical content, as illustrated by the following example.

$$\sum_{i=1}^{n-1} a^i + \frac{i+1}{i-1} \quad (1)$$

In this equation, the summation which applies to both of the remaining terms represents the first level of nesting. Two terms are contained within the summation, however, the second of these terms itself contains two complex terms in the numerator and denominator of the fraction, and this will require another level of nesting. The representation of this equation must therefore include a lengthy pause to indicate that the remaining terms are all contained within the summation, followed by additional pauses to indicate the scope of the superscript and fraction. The approximate spoken rendering of this equation would therefore be, “*the sum from i equals 1 to n minus 1 of... a to the i. plus begin fraction, i plus 1 over i minus 1, end fraction*”. Adjustments to the speaking rate are also applied at the various clause boundaries to further reinforce the segmentation of the material. In the above example, the complex terms contained within the fraction would be indicated by an increase in the speaking rate relative to the rest of the equation. However, these adjustments must be carefully

controlled to ensure that sub-expressions at deeper nesting levels are not uttered too quickly. With this in mind, a 6% increase in speaking rate was found to produce acceptable results for a number of nesting levels [15].

The prosodic model described in the above example indicates the structure of the equation in a highly intuitive manner, however clearly this approach will never be entirely sufficient, as indicated by the presence of an additional lexical cue to indicate the fraction. In this new model, lexical terms such as “*begin fraction*” are replaced with spearcons, and these are used to support the prosodic segmentation of all superscripts, subscripts and fractions. Both these spearcons and the simple, beep-like earcons which function as audio parenthesis are applied to the prosodically enhanced speech using a similar strategy which reinforces the segmentation implied with prosodic parameters such as pauses and adjustments to the speaking rate. These short beeps function as audio parenthesis in much the same way as brackets are used to delimit equations in many programming languages. An opening audio parenthesis is represented by a short beep, specifically a 20msecs sine wave pulse, which is positioned to the left of the main voice signal using binaural processing. A closing audio parenthesis is represented in the same way except the sine wave pulse is now positioned to the right. This scheme significantly enhances the structural segmentation implied by the speech prosody, and in addition, the extremely short duration and the externalizing effect of the binaural processing minimizes the detrimental effect of these additional sounds on the primary speech content. The use of spatial position to indicate opening or closing parenthesis is also very cognitively efficient as it relies on a relatively coarse and intuitive sense of spatial hearing. These two earcons were designed to support this left-right distribution as the frequency of each sine tone ramps up or down depending on whether it is an opening or closing parenthesis respectively. This frequency ramp, or glissando, is not highly apparent due to the quite short duration of the pulse, however, it does help to further distinguish the two earcons and is conceptually suggestive of the implied opening and closing delimiters. Multiple nested levels of audio parenthesis are created by adjusting both the spatial position and the frequency range of the sine wave pulse.

Parenthesis can be used to remove many of the structural ambiguities which may arise in spoken mathematics, however, the vertical layout of printed notation, with fractions, subscripts and superscripts for example, must also be indicated using some form of additional cue. In this case, the use of non-speech sounds will require the user to learn the meaning of different abstract sounds, however, the use of lexical cues is also problematic as the additional speech content may detract from the main spoken content. For this reason, spearcons based on the phrases “*frac*”, “*sup*” and “*sub*” are used to indicate fractions, superscripts and subscripts respectively. In each case, the spearcon is presented using the same left-right, binaural processing to indicate the beginning or ending of the fraction, superscript or subscript. This approach is advantageous as the original phrase provides a lexical indicator of the type of structure while the significantly reduced duration (time compressed to approx. 200msec) and simplified phrasing should also lessen the impact of the suffix effect. In addition the use of spatial positioning to indicate the beginning or end of a sub-expression further reduces the duration and detrimental effect of required lexical cues.

5 Conclusion

Working with mathematical material via audition rather than vision places an increased cognitive load on the listener. This is largely unavoidable due to critical importance of spatial structures in mathematical notation and the significant difference between audition and vision in terms of spatial resolution. The design of a system for visually impaired users must therefore concentrate on resolving structural ambiguities in a cognitively efficient manner. In addition, the system must provide the user with temporal control over the material in the form of browsing and overview capabilities, so the user can direct their own path through the material at their own pace.

In his PhD thesis on the AsTeR system, T.V. Ramen stressed the importance of identifying dimensions in the audio space to parallel the functionality of the dimensions in the visual setting [2]. This does not imply that these dimensions should be simply mapped from one modality to another, instead it is the functionality of these dimensions that should be replicated between these two modalities. The spatial layout of printed mathematical material should not therefore be simply replicated with sound. Instead the functionality of this layout must be provided via audition using a method that is appropriate for this modality. The human auditory system is particularly attuned to speech and this suggests that the most cognitively efficient means of presenting mathematical structural information via audition is through the replication of natural speech patterns, such as prosody. However, prosody alone cannot entirely resolve structural ambiguities and so additional cues are required.

While non-speech sounds such as earcons can be constructed to represent a hierarchical structure (such as nested parenthesis or menu items) this will require additional cognitive processing on the part of the listener which may only distract the processing of the mathematical material, and this is particularly true of complex schemes based upon musical structures. The audio punctuation strategy proposed in this paper overcomes this issue, as in this instance it is not the actual audio content of the earcon that is important, but rather its coarse spatial position (simply either to the left or to the right) relative to the main speech signal. In this way, the earcons function as audio parenthesis which punctuate and augments the speech content in an intuitive and non-distracting manner.

Spearcons are an interesting new development as they lie somewhere between a clearly comprehensible spoken utterance and an abstract non-speech sound. Spearcons are therefore an excellent way to indicate structural elements such as fractions, superscripts and subscripts as they are less distracting than traditional lexical cues but still provide a description of the particular structural element involved.

References

1. Gillan, D., Barraz, P., et al.: Cognitive Analysis of Equation readings: Application to the development of the MathGenie. In: Miesenberger, K., Klaus, J., Zagler, W.L., Burger, D. (eds.) ICCHP 2004. LNCS, vol. 3118, pp. 630–637. Springer, Heidelberg (2004)
2. Raman, T.V.: Audio Systems for Technical Reading. PhD Thesis, Department of Computer Science, Cornell University, Ny, USA (May 1994)

3. Gaura, P.: REMathEx – Reader and Editor of the Mathematical Expressions for Blind Students. In: Miesenberger, K., Klaus, J., Zagler, W.L. (eds.) ICCHP 2002. LNCS, vol. 2398, pp. 486–493. Springer, Heidelberg (2002)
4. Stevens, R.D.: Principles for the Design of Auditory Interfaces to Present Complex Information to Blind Computer Users. PhD Thesis, University of York, UK (1996)
5. Harling, P.A., Stevens, R., Edwards, A.: Mathgrasp: The design of an algebra manipulation tool for visually disabled mathematicians using spatial-sound and manual gestures. HCI Group, University of York, UK (1995)
6. Grantham, D.W.: Detection and discrimination of simulated motion of auditory targets in the horizontal plane. *J. Acoust. Soc. Am.* 79, 1939–1949 (1986)
7. Hollander, A.J., Furness, T.A.: Perception of Virtual Auditory Shapes. In: Proceedings of the International Conference on Auditory Displays (November 1994)
8. Goose, S., Möller, C.: A 3D Audio Only Interactive Web Browser: Using Spatialization to Convey Hypermedia Document Structure. In: Proceedings of the ACM International Conference on Multimedia, Orlando, USA, pp. 363–371 (1999)
9. Begault, D.R., Erbe, T.R.: Multi-channel spatial auditory display for speech communications. Audio Engineering Society 95th Convention, Preprint No. 3707 (1993)
10. Crispian, K., Petrie, H.: Providing Access to Graphical-Based User Interfaces for Blind People: Using Multimedia System Based on Spatial Audio Representation. 95th AES Convention, *J. Audio Eng. Soc. (Abstracts)* 41, 1060 (1993)
11. Nemeth, A.: Abraham Nemeth’s Anthology on the Unified Braille Code, <http://www.math.virginia.edu/arm4r/nemeth>
12. Baddeley, A.: *Human Memory: Theory and Practice*. Lawrence Erlbaum Associates, London (1990)
13. Walker, B., Nance, A., Lindsay, J.: Spearcons: Speech-based Earcons Improve Navigation Performance in Auditory Menus. In: Proceedings of the 12th International Conference on Auditory Display, London, UK, June 20-23 (2006)
14. Edwards, A.D.N.: Using sounds to convey complex information. In: Schick, A., Klatte, M. (eds.) *Contributions to Psychological Acoustics: Results of the Seventh Oldenburg Symposium on Psychological Acoustics*, Oldenburg, pp. 341–358 (1997)
15. Fitzpatrick, D.: Mathematics: how and what to speak. In: Miesenberger, K., Klaus, J., Zagler, W.L., Karshmer, A.I. (eds.) ICCHP 2006. LNCS, vol. 4061, pp. 1199–1206. Springer, Heidelberg (2006)

On Necessity of a New Method to Read Out Math Contents Properly in DAISY

Katsuhito Yamaguchi¹ and Masakazu Suzuki²

¹ Junior College Funabashi Campus, Nihon University
7-24-1 Narashinodai, Funabashi, Chiba 274-8501, Japan
eugene@gaea.jcn.nihon-u.ac.jp

² Institute of Systems, Information Technologies and Nanotechnologies (ISIT)
2-1-22 Momochihama, Sawara-ku, Fukuoka 814-0001, Japan
suzuki@isit.or.jp

Abstract. The necessity of a new method for defining how to read out technical terms or mathematical formulas properly in mathematical DAISY content is discussed. Two concepts of enabling aloud reading, “Ruby” and “Yomi,” are introduced. A tentative way to assign pronunciation based on “Ruby tag,” which will be adopted in the next version of DAISY, is proposed. Our new version of DAISY software, in which Ruby and Yomi functions are implemented, is presented. Combining this software with our optical character recognition system for mathematical documents allows both sighted and print-disabled people to easily produce and author mathematical DAISY books. Furthermore, people with various print disabilities can browse in mathematical content with enhanced navigation.

Keywords: DAISY, mathematics, print disability, aloud reading, inclusion.

1 Introduction

To achieve a truly inclusive society, it is clear that printed content should be more accessible to people with print disabilities. The Digital Accessible Information System (DAISY) [1], which is becoming a standard format of accessible digital books, is an excellent solution. DAISY enables people with various print disabilities such as blindness, low vision and dyslexia to access general printed content. Many documents are available as a DAISY book. For instance, the American Physical Society recently began a project aimed at making the content of “Physical Review Letters,” one of the most prestigious academic journals in physics, accessible in the DAISY format [2].

In the current version, DAISY3, mathematics is officially supported based on MathML. Some DAISY browsers such as “EasyReader” [3] and “gh PLAYER” [4] can handle the mathematical content. Combining “Save As DAISY” technology in Microsoft Word [1] with “MathType,” “MathDaisy” and “MathPlayer” [5], we can produce a DAISY mathematics book of certain quality.

However, DAISY still has problems in terms of mathematical and scientific content when it comes to user-friendliness. For instance, in the production of content, there is no way to convert a printed mathematical document directly into DAISY format. Such conversion requires certain computer skills, time and costs for sighted teachers or assistants to produce or author DAISY materials. Furthermore, people with print disabilities cannot do it for themselves.

Multilingual support other than that for European languages is another outstanding problem in DAISY. For instance, in Japanese, Chinese characters or their compounds usually have several pronunciations. In the current version of DAISY, there is no definite method to specify how to read such characters with speech synthesis. The only approach is to directly embed audio files of a human voice.

Although they are less numerous, there are similar problems that arise even in English contexts. For instance, the pronunciation of “read” differs according to its tense; if “a” represents a mathematical quantity, then it should be read as “ei;” both one gram and one unit of gravitational acceleration are written as “1g.” To treat these cases properly, approaches based on the W3C Speech Synthesis Markup Language (SSML) [6] have been studied. However, these studies do not seem to provide a perfect solution since they do not necessarily take into account multilingual mathematical content. Not only Chinese characters but also many mathematical symbols or formulas have various pronunciations. They have to be read in an appropriate manner according to their context. We also require a method to define how to read symbols with speech synthesis. SSML approaches do not seem to meet this demand. The SSML approaches aim to control pronunciation; however, in the case of mathematical formulas, we need to provide a correct word description rather than the correct pronunciation.

In this paper, after briefly discussing state-of-the-art mathematical accessibility and DAISY in section 2, we evaluate the necessity of introducing a new method of assigning how to read technical terms and mathematical formulas in section 3. In section 4, we provide a solution for the assignment and present new software for DAISY, in which our method is implemented. Other remarkable features of our software are discussed as well.

2 Mathematical Accessibility and DAISY

Recently, mathematics accessibility for visually disabled people has been gradually improved. For instance, speech-output utilities for mathematical content, automatic LaTeX-to-Braille or MathML-to-Braille converters and mathematical editors for blind people have been studied and, in some cases, put to practical use. Sophisticated software and hardware that produce tactile graphics and other interfaces have also been developed [5, 7, 8, 9, 10, 11, 12, 13, 14, 15]. Such works have greatly contributed to the production of accessible scientific material for visually disabled students.

As was reported at ICCHP 2008, for instance, our research group “Infty Project” and the non-profit organization “Science Accessibility Net” [14, 15] have

been working on the development of assisting tools for the visually disabled such as a mathematical optical character recognition (OCR) system and an accessible mathematical-document editor with speech output.

Our OCR system for mathematical documents, “InftyReader”, can properly recognize scientific documents in print or PDF, including complicated mathematical expressions, tables, graphs and other technical notations. The system converts documents into various accessible formats such as LaTeX and MathML. The mathematical-document editor “ChattyInfty” allows blind people to access mathematical and scientific documents with speech output. Information displayed in its main window is completely equivalent to that of “InftyEditor” for sighted people. All mathematical expressions are displayed in the ordinary print style. ChattyInfty reads aloud not only text but also mathematical expressions. Blind people can read, write or edit scientific documents including mathematical expressions, with speech output. Furthermore, ChattyInfty can export the “aloud reading” of scientific content as a text file, in which all the mathematical expressions are replaced with their word descriptions. Using ChattyInfty, the blind can easily share the same content with sighted classmates, colleagues or teachers in an inclusive environment. However, the current version of ChattyInfty is not necessarily useful for print-disabled people other than the blind.

DAISY has been developed by the DAISY Consortium since 1996 [1] and, among other features, was initially intended to provide a visually disabled reader random access to an audio book from a table of contents and realize a long-duration talking book in compact size. DAISY has progressed remarkably, and it can presently handle multimedia content including audio content, text and images. Text-based content is read out with speech synthesis. With its zoom function, word-tracking function and other features, DAISY allows greater access to general printed materials not only for the blind but also for the partially sighted and dyslexic. DAISY is now becoming a standard format for accessible digital books.

Whereas the previous version of DAISY XML (DAISY2) could treat a mathematical expression only as an image with an alttext (alternative text), in the current version (DAISY3), MathML can be embedded in DAISY XML format. In the alttext approach, mathematical formulas in a DAISY book were always read out continuously from the top to the end; that is, a reader could not browse within a formula. In the MathML approach, the reader can access formulas as well as text with enhanced navigation. Hence, we began a study to improve Infty software so that it can support DAISY. However, we noticed that DAISY still has problems in terms of user-friendliness when it comes to multilingual mathematical and scientific content.

3 Necessity of the Proper Aloud-Reading of Mathematical Content

In the Japanese language, four different character sets are used simultaneously in print: Chinese characters, hiragana, katakana and alphanumeric symbols. While

hiragana and katakana are essentially sets of phonetic symbols, a single Chinese character or a compound of such characters usually has several pronunciations according to context. In mathematics or science, such characters are often read in a manner that differs from the usual. For instance, the Chinese character compound for “regular” should be read as “seisoku” in mathematics but as “Masanori” when referring to a male name. Since the Chinese character for “empty” in mathematics (pronunciation: “kuu”) is also used to express “the sky” (“sora”), we cannot read it aloud without first noting the context.

We made a brief survey of the ambiguity relating to mathematical or mathematics-related technical terms that were listed by Okamoto et al. with the support of the Mathematical Society of Japan [16]. There are 10,199 entries in total. We picked all terms that had the three following characteristics: (a) a single Chinese character or a compound of characters including no hiragana or katakana, (b) a common noun, an adjective, a prefix or a suffix, (c) a word used in education ranging from elementary school to mathematics or mathematics-related courses at university. However, to avoid overlapping, we did not count compound words in principle, although some are included in the list since they are always used as one word.

There are 933 words in the intersection of categories (a), (b) and (c). Among these, 517 have no ambiguity of reading since they are manifestly mathematical or mathematics-related, and thus never used in a nontechnical context. The remaining 416 words can appear in both mathematical and nonmathematical contexts. Among these, 73.6% have only one pronunciation, whereas 26.4%, or 110 words, have more than one pronunciation according to context. Therefore, there is a clear need to specify how to read technical terms properly.

To avoid ambiguity, a Japanese printed document often includes so-called “Ruby” characters to specify how to pronounce each Chinese character or their compound. It is a kind of phonetic description typically written in hiragana or katakana to the side of the original Chinese characters in a small (Ruby-type) font. However, Ruby characters are not necessarily placed alongside all Chinese characters. In the case that there are no Ruby characters in the original printed material, the DAISY version does not include them either. It is thus impossible to determine automatically how to read characters on the basis of Ruby only. This is a serious problem for producing a Japanese-language school textbook in DAISY because the content needs to be read out even if there are no Ruby characters.

A similar situation occurs for mathematical expressions. For instance, in a mathematical formula, “X squared” sometimes should be read as “X superscript two.” We sometimes need to specify how to read characters locally according to context.

We briefly surveyed this ambiguity for all mathematical syntaxes defined in Infty software. For characters frequently used in education ranging from elementary school to mathematics or mathematics-related courses at university, the ambiguity seems to be greatest for the three following types of formulas: (a) formulas with fractions, (b) formulas with superscripts or subscripts, (c) formulas with enclosing symbols. Some examples are presented below.

a1. Fractional Type (1) $\frac{dy}{dx}$

“ dy over dx ”	a fraction
“ dy by dx ”	a derivative

a2. Fractional Type (2) $2\frac{1}{3}$

“two and one-third”	a mixed fraction
“two (times) one-third”	a product of a natural number and a fraction

b1. Super/Subscript Type (1) x^T

“ x to the T th power”	a power or an exponent
“ x transposed”	a transposed vector
“ x superscript T ”	e.g., the T th component of a vector

b2. Super/Subscript Type (2) f^{-1}

“ f to the minus-one power”	a reciprocal
“ f inverse”	an inverse function

c1. Formula with Enclosing Symbols (1) $f(x + y)$

“ f of x plus y ”	the function of $x + y$
“ f left parenthesis x plus y right parenthesis” f is a coefficient.	

c2. Formula with Enclosing Symbols (2) (a, b)

“Open interval of a and b ”	an interval
“Point a, b ”	a coordinate, a vector

This expression is used for many other entities, such as binomial coefficients and inner products.

Hence, we need a method to specify how to read the formulas properly with speech synthesis according to context.

In the forthcoming version of DAISY, DAISY4, it seems that a Ruby tag will be adopted officially in XML to specify the manner of pronunciation when Ruby characters appear explicitly in an original document. However, it is clear that this approach cannot solve the problem completely. If there are Ruby characters, a DAISY browser might be able to read the Chinese characters correctly. If not, all characters would be read out in the same manner throughout the entire document with speech synthesis. In addition, we cannot assign how to read mathematical formulas using this approach since we cannot apply Ruby characters to them.

4 “Yomi” Function and New Software for DAISY

Although the Ruby tag is expected to be adopted in the forthcoming version of DAISY, as discussed in the previous section, we need another way to define how to read out each symbol, character or mathematical formula locally besides using Ruby. We refer to this new concept of assigning a pronunciation as “Yomi” (a Japanese word that means “a manner of reading aloud”). Ruby and Yomi are conceptually different. Although Ruby is included in an original print document, Yomi is not. The reading given with Yomi, therefore, should not appear explicitly in the DAISY version, unlike the case for Ruby. However, Yomi has to control speech output like Ruby.

To introduce the Yomi function into DAISY, we tentatively present a method based on the Ruby tag. In the case that there are Ruby characters in a document, we simply use the Ruby tag, `<ruby>`, which has the following structure.

```
<ruby><rb>(an original character sequence)</rb><rt>
its word description</rt></ruby>.
```

Here, the `<rb>` tag encloses an original word, phrase or mathematical formula, and the `<rt>` tag defines how to read it. In this case, the text given in the `<rt>` tag is displayed next to the original text. In addition, we give a new attribute to the Ruby tag, “yomi.” If `type="yomi"` is declared as

```
<ruby type="yomi"><rb>(an original character sequence)</rb><rt>
its word description</rt></ruby>,
```

then the reading inside the `<rt>` tag should not appear explicitly on the display. In both cases, however, a DAISY browser should read out the text indicated in the `<rt>` tag, instead of the default.

As mentioned previously, ChattyInfty can export both the word description of a mathematical expression and its MathML form. On this basis, we worked on implementing a new function into Infty software so that it can export a recognized result or an edited file in DAISY XML format with Ruby and Yomi capabilities. We thus developed a new version of ChattyInfty to deal with mathematical DAISY content so that it is also useful for people with low vision or dyslexia. In the software, as for the previous version, a blind user can read and author mathematical documents with speech output. A low-vision reader can improve contrast or magnify the displayed content for ease of reading. Visual tracking and synchronized speech allow a dyslexic reader to maintain focus on a document. All the mathematical content is read out properly with the Ruby or Yomi function. Furthermore, since a result recognized by InftyReader can be imported directly, both the sighted and print-disabled can produce mathematical DAISY books easily from printed materials by making use of Infty software only.

The window of this new software is divided into two parts—an index window and a main window—in a similar way to other popular DAISY browsers. The

index window has a tree structure to show a table of contents in a style similar to that of Windows Explorer. This gives a user random access to a document displayed in the main window. Pressing the F6 key or Ctrl + F6, one can easily switch the active window. Since all functions in ChattyInfty for browsing and editing mathematical expressions are incorporated as a dynamic link library, the main window is completely equivalent to that in the previous version of ChattyInfty. One can not only browse but easily author mathematical DAISY content. In terms of editing Ruby and Yomi, by pressing Ctrl + Shift + Up Arrow keys after selecting part of the text or a mathematical formula, a user can access a dialog to assign it aloud-reading as either Ruby or Yomi.

In this new version, one can also export a file in DAISY3 XML format to another DAISY browser or player. Since the Ruby function is not supported in DAISY3, the reading indicated with Ruby or Yomi is replaced with a corresponding audio (wave) file. We highly expect that the Yomi function as well as the Ruby function will be added officially to the DAISY standard in the near future so that multilingual mathematical content can be dealt with properly.

5 Conclusion

We need both Ruby and Yomi functions in DAISY to read mathematical content properly. We tentatively present a new approach referred to as Yomi, which is based on the Ruby tag, and present a new version of ChattyInfty for DAISY in which Ruby and Yomi functions are implemented. Making mathematical DAISY content is not necessarily easy for teachers and assistants with ordinary sight. Using the new ChattyInfty software, they can convert printed books to accessible content even if they do not have special computer skills. People with print disabilities can not only read but also author content using the same software. We believe the software will play an important role in achieving a more inclusive environment.

Regarding graphics in mathematical DAISY content, ViewPlus and the DAISY Scalable Vector Graphics (SVG) team are presently collaborating to define authoring guidelines, define DAISY SVG attributes for improved access, and expand SVG with additional namespaces to increase capabilities in terms of both accessibility and mainstream usefulness [1, 7]. In addition, ViewPlus and our team are collaborating to improve the recognition of mathematical expressions found in graphics and to include their MathML forms (recognized results) in SVG. Combining these technologies with the new ChattyInfty software, we will be able to produce totally accessible mathematical content in the near future.

Acknowledgments. This research was supported by the Ministry of Internal Affairs and Communications, Japan, through the Strategic Information and Communications R&D Promotion Program.

References

1. The DAISY Consortium, <http://www.daisy.org/>
2. Gardner, J.A., Bulatov, V., Kelly, R.A.: Making journals accessible to the visually impaired: the future is near. *Learned Publishing* 22(4), 314–319 (2009), <http://www.viewplus.com/about/abstracts/091earnpubgardner.html>
3. Dolphin Computer Access Ltd., <http://www.yourdolphin.com/>
4. gh, LLC, <http://www.gh-accessibility.com/>
5. Design Science, Inc., <http://www.dessci.com/en/>
6. W3C, Speech Synthesis Markup Language (SSML) Version 1.0, <http://www.w3.org/TR/speech-synthesis/>, The DAISY Consortium, SSML Integration Feature, <http://www.daisy.org/z3986/2010/auth/features/ssml/0.1/z3986a-ssml.html>
7. ViewPlus Technologies, Inc., <http://www.viewplus.com/>
8. Duxbury Systems, Inc., <http://www.duxburysystems.com/>
9. Lambda Project, <http://www.lambdaproject.org/>
10. Gillan, D., Barraza, P., Karshmer, A.I., Pazuchanics, S.: Cognitive analysis of equation reading: Application to the development of the math Genie. In: Miesenberger, K., Klaus, J., Zagler, W.L., Burger, D. (eds.) ICCHP 2004. LNCS, vol. 3118, pp. 630–637. Springer, Heidelberg (2004)
11. Karshmer, A., Bledsoe, C., Stanley, P.: The Architecture of a Comprehensive Equation Browser for the Print Impaired. In: Proc. the IEEE Symposium on Visual Languages and Human Centric Computing, Dallas (2005)
12. Archambault, D., Berger, F., Moto, V.: Overview of the universal maths conversion library. In: Proc. the AAATE 2005 Conference on Assistive Technology: From Virtuality to Reality, pp. 256–260. IOS Press, Amsterdam (2005)
13. Gardner, J.A., Bulatov, V., Stowell, H.: The ViewPlus IVEO Technology for Universally Usable Graphical Information. In: Proc. the 2005 CSUN International Conference on Technology and People with Disabilities, Los Angeles (2005), <http://www.viewplus.com/about/abstracts/05csungardner2.html>
14. Yamaguchi, K., Komada, T., Kawane, F., Suzuki, M.: New Features in Math Accessibility with Infty Software. In: Miesenberger, K., Klaus, J., Zagler, W.L., Karshmer, A.I. (eds.) ICCHP 2008. LNCS, vol. 5105, pp. 893–900. Springer, Heidelberg (2008)
15. Infty Project, <http://www.inftyproject.org/en/>, Science Accessibility Net, <http://www.sciaccess.net/en/>
16. Okamoto, K., et al.: Research on Standardization of Mathematical Terms. Grant-in-Aid for Co-operative Research (A) (1993–1995), Research No.: 05306001

Innovative Methods for Accessing Mathematics by Visually Impaired Users

Giuseppe Nicotra¹, Giovanni Bertoni², Enrico Bortolazzi², and Luciana Formenti²

¹ Dipartimento di Informatica e Comunicazione, Università degli Studi di Milano,
via Comelico 39/41, Milano, 20135, Italy
nicotra@dico.unimi.it

² Veia Progetti s.r.l, via Casorati 2 Verona, Italy
{bertoni, Bortolazzi}@veia.it, lucy@lambdaproject.org

Abstract. Blind people have always considered the study of Mathematics as a difficult problem, hindering the chance to approach scientific studies for generations of visually impaired people. At present computer use is very widespread among blind students, who appreciate more and more its advantages (speed, efficiency, access to almost limitless papers), yet in the field of Mathematics, there are still few benefits, due to its complex symbols and because notation is graphical, whereas Braille notation is linear, to suit characteristics of the sense of touch.

The LAMBDA-project team designed a system based on the functional integration of a linear mathematical code and an editor for the visualization, writing and manipulation of texts.

Keywords: Braille, Mathematics, blind people.

1 Introduction

The study of mathematics has been always considered as an uneasily solvable problem for the blind and has strongly influenced the access to scientific studies by whole generations of visually impaired people.

Today, computer use is very widespread among blind students, who have increasingly appreciated its benefits in studying maths, thanks to such recently developed programs as Lambda (www.lambdaproject.org). In the field of mathematics transcriptions, however, due to its complex symbols and non-linear writing, expert manual intervention is still necessary. In fact, unlike from other text forms, experts cannot take advantage from OCRs and automatic Braille recognition and conversion systems.

2 Teaching Maths to Blind Children and Students

Besides being a laboratory for innovations for teachers, operators and parents, teaching maths to blind children and students has long been carefully studied by typhlogists and pedagogists. [1]

Replacing the information which normally comes from the visual canal with tactile or verbal information is not always easy; yet, convincing teaching methods have developed in recent years, with a wide range of tools of various kind, especially manipulators, which can support visually impaired students in the various stages of learning mathematics.

At present, we are experiencing a stage of transformation, rich in potentials and featured especially by the following two aspects:

- progressive and continuous perfecting of the model of integrated schools;
- introduction of computer tools even among the youngest.

Even though they are not bound by causal factors, both aspects are in strict connection, since computers can be useful to overcome the steady need of specialised operators, while consequently favouring the integration of blind pupils in conventional schools.

For the youngest pupils, conventional writing tools remain irreplaceable, especially for the need of checking the text on more lines, while also catching the vertical structure (for instance, with column operations).

Using Braille displays to do maths requires a good knowledge of the tool and reliable training. First of all, it needs the ability to mentally build the bidimensional image of formulas and/or expressions, starting from the information offered line by line.

Of course, the use of computers will be always supported by raised graphics, three-dimensional models, manipulation and exploration teaching aids, and any others developed in years of typhlo-teaching experience.

Even though less immediate and effective, PCs with voice synthesis are used to read and manage maths texts, too

3 Problems Related to Maths Text Transcription

Managing maths documents has long been a difficult problem for blind students.

Such programs as Lambda have allowed to:

- offer blind students the possibility to perform maths exercises with PCs, rapidly and efficiently, while ensuring teachers, particularly those of integrated schools who do not know any Braille, direct control over learning processes at the same time;
- develop an efficient, powerful and flexible maths editor for high school and university students, which can convert both input and output documents into MathML standard;
- offer transcription operators, teachers and parents efficient tools to cut the times of preparing Braille maths texts. It must be mentioned that, unfortunately, this operation still needs manual text insertion. Therefore, operators must be specifically skilled in both maths and Braille maths.

Complaints and dissatisfaction from transcription centres about delays in producing Braille school digital texts, especially maths texts, are not rarely reported on forums dedicated to school and school integration and at parent meetings organised in the various local sections. In fact, operations on Braille texts, unlike other schoolbooks, cannot be automatically processed through OCR, because the information obtained cannot be used for Braille conversion.

4 Braille National Codes and the Search for a Maths Standard. LAMBDA's New Resources

Despite its worldwide popularity, the Braille system allows univocal codification of alphabetical characters and of few punctuation signs only. The differences among the various countries are often due to the specific needs of the relevant language. Therefore, they have a logical explanation. In some cases, however, they merely result from the lack of contacts and agreements.

Unfortunately, precisely such universal languages as music and maths have been more strongly affected by this lack of communication and have originated the widest differences among the various nations. Obviously, these differences have reduced the possibility to exchange Braille documents worldwide. In the music field, the lack was partly remedied by an international Braille manual of music signs, which was recognised by 16 countries. Yet, it was not published earlier than in 1996.

Information technology has hugely reduced the hindrances to communication caused by the many sophisticated Braille signs, even though the new 8-dot code hasn't been standardized at the international level, either (there are at least ten different Braille codes in Europe).

These differences have a more modest weight, however, because they influence end representation, as decided by each single user, while the document remains the same. A French visually impaired user can write a text on the PC and read it according to its Braille sets of signs; then send it to an Italian user who will read the same text visualised according the Braille code in use in Italy. It is clear that all this could have not been possible if documents had been exchanged on paper instead of electronic format.

Things change with non-textual languages such as maths.

In this case, the differences among national codes do not limit themselves to the combination of Braille dots used to represent the single characters, but also extend to the rules used for transforming graphical and bidimensional information into a linear sequence.

The problem does not simply concern blind people, but even all those who must use textual systems to transfer information of this kind. Maths experts have long solved this problem with the help of systems, such as TeX and LaTeX[2], which separate the textual description of a formula (compatible with various platforms and easily transferable and exportable) from the end graphical representation obtained by filling in the text code, with the use of a specific software program.

The TeX system, which provides powerful customisation tools, is being continuously enriched and perfected by single authors, with more and more sophisticated results, against a reduction in the exchange potential of the code, which, in addition, does not easily adapt to the purpose of automatic conversion processes.

This is very different for MathML. This XML code, which was defined by a work group of W3C, was born to put maths formulas into the Web, but is now gaining a primary role as an interchange and transformation tool among maths documents in different formats, thanks to the strict specification requirements and reliability ensured by the W3C (the web consortium)

The MathML standard is quite sophisticated. It develops itself on a very high number of lines, with very strict validation limits.

A valid solution has been recently offered by the LAMBDA project with a system based on functional integration of a linear maths code and an editor for visualisation, writing and manipulation.

The **code** (Lambda Math Code) is a direct derivation of MathML, which was developed for optimal use with Braille peripherals and voice synthesis and real time automatic univocal conversion into an equivalent input/output MathML version.

The **editor** allows linear writing and manipulation of maths expressions and offers a series of *compensation* functions, that is a number of help options to reduce the user's understanding and managing difficulties caused by the visual impairment and the consequent need of a linear code for managing formulas. LAMBDA has been developed for students ranging from first-grade secondary schools to universities; some reasonable skill in using PCs is needed. (www.lambdaproject.org)

The Lambda code will overcome the problems related to different Braille representation in the various languages, because its structure and maths text organisation is univocal. The only change is its 8-dot Braille representation.

The Lambda code is of free use for both users and developers. The ownership of the LAMBDA format belongs to the European Blind Union (EBU), which supervises the steady development of the Lambda format, besides promoting its use throughout Europe.

Lambda is a linear, text-format, efficient, full and univocal maths code, also easily manageable with Braille display or voice synthesis through the low-cost dedicated software developed by the LAMBDA project.

Being essentially different in formatting (linear against multi-line) and tag customisation (more or less expanded, explicit, in line with user's needs and preferences), the Lambda code can maintain full biunivocal correspondence with the structural elements of MathML.

Each marker has been assigned:

- an 8-dot Braille combination, made up of one or more characters;
- a text description that must be read with synthesis;
- a graphical representation (font) for visualisation on the screen.

The 8-dot Braille LAMBDA code offer users best clarity, functionality and essentiality, by keeping the widest possible analogies with the equivalent 6-dot Braille. With Lambda, it is also possible to obtain the corresponding 6-dot Italian Braille code for print.

Lambda has been also equipped with a video representation that allows any sighted, sufficiently informed user to understand the meaning of the text included, even when he/she has no knowledge of Braille signs and the text has been linearly organized.

Moreover, Lambda has been completed with a clear and efficient font True Type, by giving a special symbol to each marker.

5 LAMBDA Editor Applications for Studying Mathematics

The main advantages offered by the Lambda editor for non-sighted pupils in the maths field can be summarized as follows:

- best efficiency in performing school works, with reduction of transcription and manipulation times;
- unlike conventional Braille, the maths texts written on PCs can be understood even by those who do not know the code; this is particularly important for relations with maths teachers in integrated schools (or universities) who can interact in the teaching process, directly and reliably;
- possibilities of contacts and exchanges of information among classmates;
- access to maths documents written in standard digital format (for instance MathML);
- access to maths documents produced by teachers with their own programs and easily convertible into MathML, automatically, for instance Mathematica, MathType, MS Word Equation Editor (the latter in combination with Math Type).
- possibility to make use of browsing, manipulation, control tools which can help non-sighted pupils by compensating, partially at least, the lack of information and management possibilities caused by visual impairment.

6 The INFITY OCR

The INFITY Project has been developed by a research organisation made up of more universities and institutes and coordinated by Japan's Kyushu University.

The INFITY reader has become a commercial product. It consists of an OCR system for maths documents.

It operates through four procedures: format analysis, conventional character recognition, structural analysis of maths expressions, manual correction of possible recognition mistakes.

The software recognises either b/w pages acquired from scanner or TIFF, GIF and 600 DPI/400 DPI PNG images.

Output of INFITY Reader recognised texts takes place in various formats: IML, LaTeX, HR-TeX, XHTML.

IML is a proprietary file made for InfyEditor, which is a graphical maths editor.

The InfyEditor allows correcting and editing recognition results, which is made easier by double representation of the original image format and recognised maths-text.

HR-TeX is a simplified version of LaTeX, which has introduced an easier form of reading maths texts and could be accessed for reading by particularly skilled non-sighted students, such as university ones.

The XHTML format contains both HTML text and maths expressions. The latter have been filed using MathML notation.

INFITY also allows direct recognition and conversion of PDF files, namely the format that editors more frequently supply transcription centres with.

Some tests have proved that the recognition level is very high (practically no-error) for textbooks of first-grade secondary schools and very appreciable for more complicated expressions of second-grade ones.

Therefore, INFITY is quite suitable for merging with a more sophisticated system that allows automatic transcription of Braille maths documents.

7 INFTY2BRAILLE: The End Product

INFTY2BRAILLE is one of the projects promoted for "New Technologies and Disabilities", Action 6, and funded by Italy's Ministry of Education. The promoting partner of the project was the Liceo Scientifico Statale N. Copernico, Bologna.

The INFTY2BRAILLE project was meant to meet this positive maths teaching experience with best exploitation of the potentials offered by PCs and by the maths OCR INFTY, in order to ease acquisition of maths texts, conversion into Lambda and 6-dot Braille, for those who need it, and video and/or Braille line representation.

The INFTY2BRAILLE project is a significant help for teachers, readers, parents of non-sighted students in secondary schools and universities and transcription centres for automatic acquisition of maths books. The INFTY OCR and INFTY2BRAILLE program will make it easy and possible to reproduce single pages of textbooks into Braille and LAMBDA formats, for the purpose of studying or performing exercises and class works. The project is addressed to students of different ages and skills (from first-grade secondary schools to university).

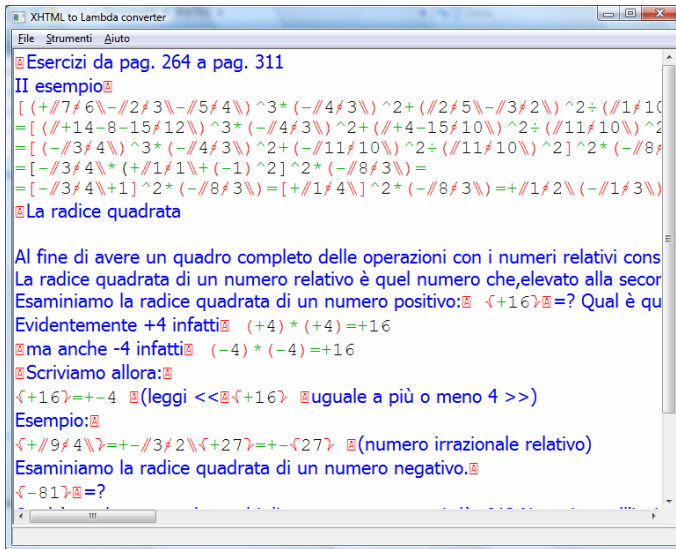


Fig. 2. Result in Lambda editor, of a page of a math book used in Italian schools

8 How Does INFTY2 Braille Work

The process of electronic maths text production can start from either 600-dot scanning of a paper book, or converting PDF pages into TIFF (preferably 600 dpi resolution).

Some basic processing is needed before converting the image. It principally consists of deleting drawings, coloured boxes, as well as any other objects in the page, which make recognition more difficult. Sometimes it is also necessary to modify the background, for instance coloured and gradient backgrounds and text colours. This

'cleaning' operation can be performed with an image processing software. A specific script for GIMP, a free software for image processing, was written to speed this process up. The script automatically uses a series of graphical filters, particularly on colours, and makes the image ready for conversion with INFTY.

Once the conversion with INFTY is made, it shows the result with InftyEditor for any further changes or corrections. In fact, after a recognition stage has taken place, it is usually needed to:

- check the page layout and delete any images left;
- check textual and maths parts;
- check that no textual characters have been incorrectly recognised;
- check that maths formulas are correct.

This stage ends up by exporting the correct file into XHTML.

The Lambda program has been expanded with a new function, which allows opening the previously saved XHTML file and visualise it with the Lambda editor.

Conversion can be checked within the Lambda program, too. At the end of the process, the document must be reformatted, so as to ease writing, for instance with intervention on too long text lines. Other problems related to formatting are due to the conversion of two-column pages, or maths expressions with results written on a separate column. In these cases, it is needed to check that the flux of reading is coherent and easy, considering that non-sighted students read one line at a time on a Braille line, which usually visualises 40 characters per line.

Another step consists of replacing the images deleted before conversion with another text, if necessary, so as to avoid losing the associated information.

9 Tests

The above-mentioned process has been perfected and tested with different text pages throughout the INFTY2BRAILLE project duration. The tests made by Braille transcription experts, teachers and computer experts have highlighted the following aspects:

- as expected, transcription is not completely automatic. Instead, it requires operators skilled both from the technical viewpoint, especially in the first stage – preparation of images for recognition - and the teaching viewpoint, especially in the end stage -document formatting-;
- learning how to use the tools (INFTY, GIMP, Lambda) is quite easy and does not require any specific computer skills;
- transcription times are largely influenced by the original format, either texts on paper for the scanner or PDF files, type of page, quantity of text and maths and their ratio. Yet, they are largely lower than the times usually needed, up to reductions of over 60%.

10 Braillelab Service

The Infty2Braille project has allowed launching a Braille transcription service for maths books of first and second-grade secondary schools, exploiting the potentials of

new technologies to solve the problems of accessing maths schoolbooks by non-sighted students. The service was started at the headquarters of the "Braille LAB" association in Verona.

The books are supplied in electronic Lambda format, 8-dot Braille, for immediate use with Lambda maths editor. The latter is provided to students and schools along with the first book requested.

Further graphical information, such as geometric images of function graphs, will be so described as to allow teachers and study assistants to arrange raised images with gummed tops or raised sheets obtained with a "heater". The web site of the association is www.braillelab.org

11 Conclusion

As an innovative product in the educational domain, INFTY2BRAILLE involves and promotes as well a change in the teaching method, e.g. to produce math text books for visually impaired students. INFTY2BRAILLE proves, once again, that mainstream schooling really includes disabled students when it uses a common language, more powerful than the traditional one, since it is a structured language (e.g. LaTeX or any HTML language). This new language is able to reduce communication hindrances within a classroom, to produce more traceable documents, less ambiguous than handwriting or photocopies. Then inclusion becomes a daily practice, intertwined with educational programming involving several different needs, such as individual teaching, and able to reach and involve all classmates and each single student.

As to the positive influence on the domain of personal relationships among classmates, the introduction of INFTY2BRAILLE builds more on the programming and on the classmates fellowship than on the tools to be used. A very important aspect is the sharing of different ways used by students to solve the same problem that challenges the whole class (teacher and students) to elaborate problem solving strategies that really makes a common richness out of an individual diversity. -

References

1. Del Campo, J.E.F.: La enseñanza de la matematica a los ciegos, Madrid: Organización Nacional de Ciegos Españoles, Sección de Educación D.L. (1996/2000); Fogarolo F.: Software per la matematica con periferiche Braille Atti convegno Handimatica, Bologna (1998); Von Prondzinsky S.: Premesse psicopedagogiche e presupposti didattici per l'uso degli strumenti informatici nelle attività logico-matematiche Atti convegno Handimatica, Bologna (1998); Quatraro, A.: Tecnologia e integrazione dei disabili visivi, BIC, Monza (2001); Graziani, P.: Le ricerche sugli ausili tiflo-informatici per lo studio della matematica Atti Convegno Handimatica, Bologna (1998)
2. Artico, G.: Il TeX e i non vedenti Università di Padova (1998)

ICCHP Keynote: Designing Haptic Interaction for a Collaborative World

Gerhard Weber

Human Computer Interaction Research Group
Technical University of Dresden
Nöthnitzer Straße 46, 01187 Dresden, Germany
gerhard.weber@tu-dresden.de

Abstract. The design of haptic interaction for blind users of a large tactile display capable of multitouch input may face a design barrier. Visual media and their modalities have to be mapped to tactile graphics and audiohaptic modalities due to the low resolution. We present the work of the Hyperbraille project on understanding the limitations of current screenreaders, implementation of a modular Braille window system merging multiple tactile views through Add-Ins, and report about the main findings of some of the evaluations related to reading Braille and gestural input.

Keywords: tactile display, screen reader, audio-haptic interaction, gestures.

1 Introduction

Collaboration among blind and sighted computer users has become possible through screenreaders. This assistive technology turns a graphical or multimedia user interface into a multimodal system supporting speech and braille output while allowing both keyboard and mouse input. Blind and sighted users read, edit or browse nearly the same documents irrespective of their origin. Blind and sighted users may even operate applications at the same time, sharing keyboard and non-visual output as well as mouse and visual output. Collaboration of this type ensures integration of blind people in education, at the work place and even for many leisure time activities as long as screenreaders continue to match the features of the visual with a non-visual user interface and vice versa.

Current commercial screenreaders are very well capable of rendering textual data. If screenreaders fail, possibly due to lack of textual data, other means of communication or other types of services may exist but more than once it has led to exclusion of blind people from daily activities.

This paper describes the efforts in the Hyperbraille project to overcome some of the limitations of current screenreaders through a large tactile display supporting multitouch input and introduces to a Special Thematic Session covering this and similar research. We discuss various designs of non-verbal interaction for handling notations and tactile views of textual and graphical contents, in particular based on novel tactile displays. In particular we show how designing, implementing and evaluating a screenreader for a large planar tactile display can be performed.

2 Large Tactile Displays

Only few changes were applied to the basic interaction design of screen readers for GUIs since their development in the 1990s. Synthetic speech actively approaches the user in return to keyboard input and describes the focus. Braille is presented such that the focus can be tracked with one's finger, often simplified by arranging the Braille on the left side of a Braille display. Other than speech can Braille displays allow routing: a user clicks on the focus that is on a menu option, a button, or a character. This interaction is close to mouse clicks but more complex mouse interaction such as drag and drop operation requires a modal technique and is hence large impractical.

The advent of large tactile displays and appropriate input devices may change the approach to interaction via screenreaders

- by improving the possibilities for exploring graphical contents or textual contents whose layout is to be preserved tactile,
- by controlling the interaction through a single point of contact, or
- by multitouch operation of haptic objects more complex than Braille or tactions.

Early systems have shown there is potential for improving the overview and hence less misinterpretation about the status of the dialog. The very first prototypical Braille window system for a planar tactile display has been proposed for a Personal Information Manager application (PIM) and included a calendar tool, an editor and a messaging service [2]. Although not comparable to the functionality of a screenreader, this user interface has shown the opportunities of bi-manual interaction on a large tactile display with 120 by 60 pins. For example, users may rely on their kinesthetic sense when entering annotations for calendar items and more easily explore the tabular arranged dates with one hand and compare a note for that date with the second hand.

Limitations of this PIM were considerable: no window frame was introduced, only one "private dialog" for command lines was used, no scrolling was possible, and no mouse input has been possible. Frequent moves between keyboard and the extra time needed for re-establishing the context on the tactile display slows down speed of interaction in this setup considerably. Other approaches with different sizes and resolution of the tactile display are not affecting the nature of this interaction as the main intention is the provision of graphical contents only [1].

Fingers should stay on the tactile display as long as possible and in particular when exploring the focus and its context. A single touch gesture recognition system allowing users to actively read and control the presentation without leaving the large tactile display prevents frequent changes [19]. Moreover, active reading includes the notion of taking the reading movement into account and separating it from gestures when determining feedback for the user.

Operating the tactile user interface by tapping within tactile graphics on links has been demonstrated for browsing hypertext documents [3]. In this scenario tactile images become navigable through tapping on tactile symbols by a finger. The use of a pen as an input device for drawing by blind people is discussed in a newer scenario using a different large tactile display [8].

More complex haptic interaction requires an input technology for a large tactile display based on multitouch principles, its input resolution is preferably superior to its output resolution. It should support also recognition of fast hand movements. These

features are similar to multitouch input for visual displays, except that haptic output may guide the fingers in a more immediate manner.

2.1 User Centered Design for Haptic Interaction

When designing haptic interaction for non-collaborative use the designer may be blind himself. Braille was developed in this way by establishing reading and writing skills at the same time. As Louis Braille was an assistant teacher at the Paris School for the Blind he could evaluate his system with a considerable number of blind users.

Designing a first screenreader for a large tactile display has also been the work of a blind person and has led to TAWIS [6]. TAWIS is a pilot software system for demonstrating the usefulness of a pin-matrix device and has grown over nearly a decade. It has many features including a window system and image processing functions allowing the user to extract outlines and to filter colors. The state of TAWIS may be not always be inspected by sighted people very easily (see Fig 1).

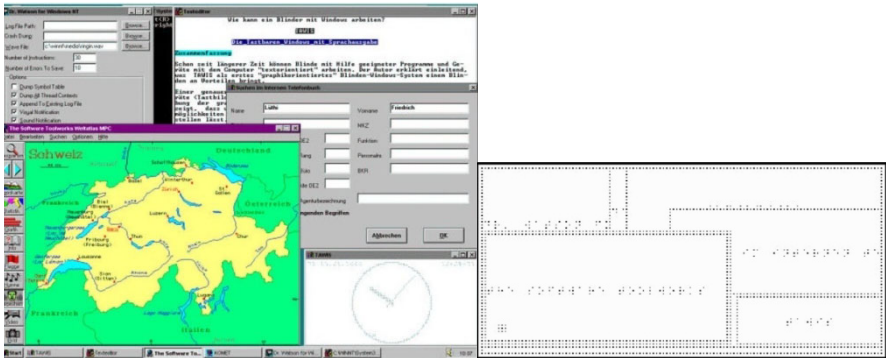


Fig. 1. A GUI multi-window setup and corresponding TAWIS Braille windows

In order to develop a screenreader for collaborative use of graphical user interfaces the design of haptic interaction is considered far more difficult than for verbal or verbalized elements of the dialog. Such a design has to enable collaboration and is itself the result of collaboration among sighted and a considerable number of blind people. We refer to this as the design barrier of haptic interaction.

Due to lack of experience with the particular haptic technologies and lack of understanding of the processes needed to cope with the non-verbal nature of haptics it appears for such a complex team necessary to rely on a user centered design process as an initial effort. UCD is an iterative process consisting of an analysis of the context of use, the description of users, design artifacts in various fidelities, and evaluation. However discrepancies between the modality analysed and the modalities used for reporting the analysis may require additional and actually unwanted iterations. The Hyperbraille project team tries carefully to understand the pitfalls arising from mismatches in the communication between sighted and blind people at each stage of the development process.

2.2 Requirements Analysis

It took a considerable effort for most blind people to master their screenreader and only few users can call themselves experts. A new screenreader should be modular enough to support existing screenreaders. The Hyperbraille project adds to conventional screenreaders additional information (both spoken and tactile) and more interaction features. We refer to this type of assistive technology as a screenexplorer in general, within the Hyperbraille project this software is called Hyperreader. Only specific applications take advantage of the spatial tactile layout of haptic characters, textures, and drawings. Through a tactile window system a conventional Braille display may be emulated, including the ability for routing.

The analysis of drawbacks of specific applications such as browser, word processor, or spreadsheet application is possible by mock-ups. A certain state of a dialog (called a scenario) which is seen by the scenario authors as typical or important for re-design is designed with a drawing program using its GUI and printed as tactile graphics. An evaluation with both expert screenreader users and beginning to intermediate users in separate groups can reveal additional insights [7]. The results of this study indicate some of the mock-ups were confirmed by blind participants to be well-designed. Other mock-ups failed and were rejected due to unclear reasons. The authors of a mock-up may be blind or sighted as the drawing program used was accessible to both kind of users. Both kinds of authors were in particular interested in scenarios requiring an overview. For example, one scenario considered a dialogue box opened and accessible at the same time as a word processor application window such as for spelling correction. Another scenario considered a word processor with additional information on text style in a separate window. Classical screenreaders cannot handle multiple windows and restrict themselves to handle the foreground window only. As a result context of spelling errors is difficult to recognize.

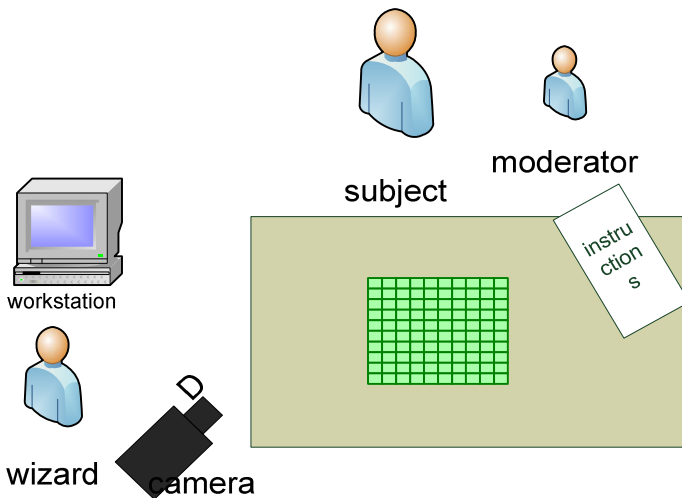


Fig. 2. Wizard-of Oz Setup for tactile Scenarios

A wizard-of-oz study with a re-design of some of the mock-ups showed the need for a Braille window system with tiled window management. The number of Braille windows presented at the same time is to be limited [9]. Unlike through paper mock-ups was this study conducted with the real tactile display and allowed to “surprise” blind users with new tactile images as a result of their input as interpreted by the wizard (see Fig. 2). The design of Braille windows for Hyperbraille is discussed in more detail by [11]. In particular acceptance of multiple tactile views is discussed by the authors. Four degrees of abstractions from the visual presentation are proposed and include layout, outline, symbol and operating view, where just the operating view is Braille-only.

Requirements for input operations by users are difficult to determine without relying on a prototype system. Although gestures on a large tactile display have been studied before [19], two-handed gestures and arbitrary gesture shapes are little understood. Sturm describes the use of finger painting colors to analyze such gestural input [15]. The above mentioned wizard-of-oz study uses the wizard’s ability to recognize gestures. All of these studies use a fixed set of gestures, for example to avoid tapping or keyboard operation but resize a Braille window or scrolling and zooming operations. However in interviews some blind participants are strongly interested in creating their own gestures to speed up the interaction. Schmidt 13 describes the ability to support learning of gestures through a specific gesture training program that is guiding the blind user’s fingers on the Hyperbraille display. Learning more complex gestures may be necessary in order to counteract flexibility of multitouch gestures and is discussed in this STS [12].

These studies also confirm the need to locate the standard keyboard as well as function keys and Braille input keys as close to the tactile display surface as possible. Moreover, the design of the hardware casing should consider ergonomic requirements as the height of the display cannot be neglected.

Making all of the requirements accessible to software and hardware developers is done through reports including text, tables and images showing shapes of widgets. It has turned out, understanding of the requirements by developers requires monitoring by blind people through functional testing of the software achievements and the continuous development of a manual. More than once are developers applying mouse input in their functional tests and do not detect the issues related to keyboard operations. An agile process model such as SCRUM requires participation of blind users on a frequent basis as long as the management of ‘sprints’ is well communicated within the team.

2.3 Implementation

The hardware of the Hyperbraille display consists of is described by Völkel et.al. [18] in detail. The software architecture is based around a hierarchical off-screen model (OSM) [5]. This off-screen model captures filtered data and computes feedback for user input. As a multimodal system, Hyperreader generates both for mouse and keyboard input appropriate spoken output or shows an update on the tactile display. Key to the user will be support for clicking on haptic objects. The off-screen model allows computing from touch input the locations on the screen that are actually referred to by the finger.

During informal tests blind user showed a strong interest in merging some aspects of the above outlined four different tactile views (see Fig. 3).

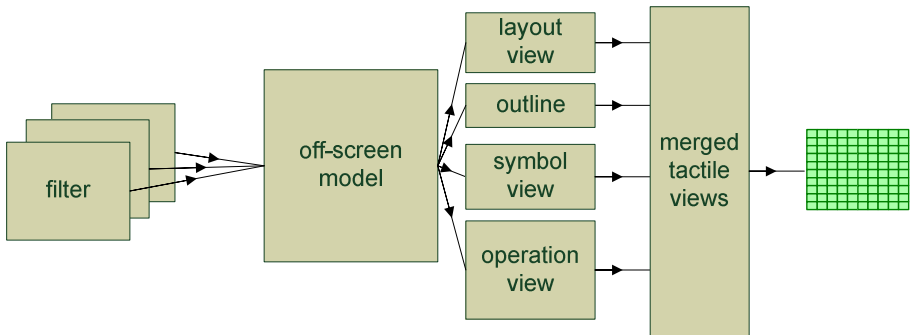


Fig. 3. Multiple tactile views are merged by Hyperreader possibly for each widget

Spindler et.al. [14] describe a new approach to modularize a screenreader by relying on Add-ins. Add-ins are a process-communication technology based on DotNet allowing changes to software applications at runtime. This ensures filters dealing with UI Automation data can be replaced by filters dealing with the DOM of a browser application, for example. Add-ins are also used for rendering of output in very different ways. Köhlmann describes the requirements for an Add-in more suitable for word processing [4]. In another study, Hyperreader allows zooming below the size of one pixel in order to provide the blind user with a haptic texture resembling the colors of each pixel but adding tactile shapes for distinguishing them [16]. Selecting names of colors allows users to select foreground and background colors according to their knowledge of the application. Finally, Add-ins allow for flexible integration of many sources of input, in particular gestures in combination with regular keyboard input.

Hyperreader may even provide a different approach to assistive technology for visually impaired people by reducing the number of windows open at a time while using large fonts and good color contrast [17].

2.4 Evaluation

The last phase in each iteration of UCD deals with evaluations involving end users. The Hyperbraille project conducts several studies related to novel features, in particular reading Braille on a large display [10], the use of Braille windows, multiple tactile views and gestures [13], the interaction with symbolic views on widgets, or the ability to learn gestures. These are controlled experiments allowing a comparison of the results of users in order to analyze the usability of the user interface.

At this point we can already conclude reading equidistant Braille is similar to Braille on regular displays. A short training period for well-experienced Braille reader is necessary to achieve the same efficiency. Beginners will possibly benefit from the slighted enlarged distance among Braille characters.

3 Outlook

Focus of future work is on applications which utilize the interactive features of bi-manual input. For example a drawing program such as Visio or Powerpoint will allow

editing of simple graphical input and drag and drop techniques may allow drawing boxes, lines, etc..

Some applications have been designed within Hyperbraille to utilize the interactive generation of tactile graphics for exploring specific data sets. A map application provides access to OpenStreetMap GIS data at zoom levels appropriate to pedestrians [20] and will be the basis for future mobile scenarios.

Acknowledgments. The HyperBraille project is sponsored by the German Ministry of Economy and Technology (BMW) under the grant number 01MT07004. Only the authors of this paper are responsible for its content. The Consortium consists of Metec AG, F.H. Papenmeier, Human-Computer Interaction Research Group at the Technical University of Dresden, Institute for Visualization, IZFM institute and IMS-Chips at the University of Stuttgart, the Research Group for Information Processing and Communication at the University of Potsdam and the Blista Marburg as well as T.O.P. GmbH & Co. KG Freiburg.

References

1. Bruder, I., Jaworek, G.: Blind and Visually Impaired People: Human-Computer Interaction and Access to Graphics. In: Miesenberger, K., Klaus, J., Zagler, W.L., Karshmer, A.I. (eds.) ICCHP 2008. LNCS, vol. 5105, pp. 767–769. Springer, Heidelberg (2008)
2. Klöpfer, K.: Ein multifunktionaler Büroarbeitsplatz für Blinde (A multifunctional work place for the Blind). Dissertation, Universität Stuttgart, Institut für Informatik (1987)
3. Kochanek, D.: A Hypertext system for the blind newspaper reader. In: Zagler, W. (ed.) Computers for Handicapped Persons, pp. 285–293 (1992)
4. Köhlmann, W., Zinke, F., Schiewe, M., Jürgensen, H.: User-Interface Filter for Two-Dimensional Haptic Interaction. In: Miesenberger, K. (ed.) ICCHP 2010, Part II. LNCS, vol. 6180, pp. 502–509. Springer, Heidelberg (2010)
5. Kraus, M., Völkel, T., Weber, G.: An Off-Screen Model for Tactile Graphical User Interfaces. In: Miesenberger, K., Klaus, J., Zagler, W.L., Karshmer, A.I. (eds.) ICCHP 2008. LNCS, vol. 5105, pp. 865–872. Springer, Heidelberg (2008)
6. Lüthi, F.: Personal communication, September 28 (2006)
7. Miao, M., Köhlmann, W., Schiewe, M., Weber, G.: Tactile Paper Prototyping with Blind Subjects. In: Altinsoy, M.E., Jekosch, U., Brewster, S. (eds.) HAID 2009. LNCS, vol. 5763, pp. 81–90. Springer, Heidelberg (2009)
8. Nishi, A., Fukuda, R.: Graphic Editor for Visually Impaired Users. In: Miesenberger, K., Klaus, J., Zagler, W.L., Karshmer, A.I. (eds.) ICCHP 2006. LNCS, vol. 4061, pp. 1139–1146. Springer, Heidelberg (2006)
9. Prescher, D.: A tactile window system for multitouch operation, Diplomathesis, Dept. Computer Science, TU Dresden, unpubl (2008)
10. Prescher, D., Nadig, O., Weber, G.: Reading Braille and Tactile Ink-print on a Planar Tactile Display. In: Miesenberger, K., et al. (eds.) ICCHP 2010. LNCS, vol. 6180, pp. 486–493. Springer, Heidelberg (2010)
11. Schiewe, M., Köhlmann, W., Nadig, O., Weber, G.: What You Feel Is What You Get: Mapping GUIs on Planar Tactile Displays. In: Stephanidis, C. (ed.) UAHCI 2009, Held as Part of HCI International 2009, San Diego, CA, USA, July 19–24. LNCS, vol. 5615, pp. 564–573. Springer, Heidelberg (2009)

12. Schmidt, M., Weber, G.: Enhancing Single Stroke Gesture Classifiers to Multitouch support. In: Miesenberger, K., et al. (eds.) ICCHP 2010, Part II. LNCS, vol. 6180, pp. 494–501. Springer, Heidelberg (2010)
13. Schmidt, M., Weber, G.: Multitouch Haptic Interaction. In: Stephanidis, C. (ed.) Universal Access in Human-Computer Interaction. Applications and Services 5th International Conference, UAHCI 2009, Held as Part of HCI International 2009, San Diego, CA, USA. LNCS, vol. 5615, pp. 574–582. Springer, Heidelberg (2009)
14. Spindler, M., Kraus, M., Weber, G.: A Graphical Tactile Screen-Explorer. In: Miesenberger, K., et al. (eds.) ICCHP 2010, Part II. LNCS, vol. 6180, pp. 478–485. Springer, Heidelberg (2010)
15. Sturm, I., Köhlmann, W., Schiewe, M., Jürgensen, H.: Communicating through gestures without visual feedback. In: Makedon, F., Maglogiannis, I., Baillie, L., Pantziou, G., Karkaletsis, V. (eds.) Proceedings of PETRA 2009, Corfu, Greece, June 9–13, p. 8. ACM Press, New York (2009)
16. Taras, C., Ertl, T.: Interaction with Colored Graphical Representations on Braille Devices. In: Stephanidis, C. (ed.) Universal Access in Human-Computer Interaction. Applications and Services 5th International Conference, UAHCI 2009. LNCS, vol. 5615, pp. 164–173. Springer, Heidelberg (2009)
17. Taras, C., Raschke, M., Schlegel, T., Ertl, T., Prescher, D., Weber, G.: Improving Screen Magnification Using the HyperBraille Multiview Windowing Technique. In: Miesenberger, K., et al. (eds.) ICCHP 2010, Part II. LNCS, vol. 6180, pp. 510–516. Springer, Heidelberg (2010)
18. Völkel, T., Weber, G., Baumann, U.: Tactile Graphics revised: The novel BrailleDis 9000 Pin-Matrix Device with Multitouch Input. In: Miesenberger, K., Klaus, J., Zagler, W.L., Karshmer, A.I. (eds.) ICCHP 2008. LNCS, vol. 5105, pp. 835–842. Springer, Heidelberg (2008)
19. Weber, G.: Gestures as a means for the blind to interact with a computer. In: Bullinger, H.-J. (ed.) Human-Computer Interaction INTERACT 1987, pp. 593–595. North Holland, Amsterdam (1987)
20. Zeng, L., Weber, G.: Audio-haptic Browser for a Geographical Information System. In: Miesenberger, K., et al. (eds.) ICCHP 2010, Part II. LNCS, vol. 6180, pp. 470–477. Springer, Heidelberg (2010)

Improving the Accessibility of ASCII Graphics for the Blind Students

Producing My Own Graphics

Karin Müller and Angela Constantinescu

Karlsruhe Institute of Technology (KIT)
Study Center for the Visually Impaired Students (SZS)
Engesserstraße 4
D-76131 Karlsruhe
{karin.e.mueller,angela.constantinescu}@kit.edu
<http://www.szs.kit.edu>

Abstract. Graphics are pervasively used in natural sciences. Making graphics accessible to blind people is difficult because blind students have to know how the layout looks like, to understand what the layout means and to have a possibility to create a graphic by themselves. Therefore we suggest to use ASCII graphics. Our approach tries to improve the accessibility of ASCII graphics by automatically extracting and transforming them in order to be printed with a tactile embosser.

Keywords: accessibility of graphics, ASCII graphics, blind.

1 Introduction

Graphics in general are not accessible to blind people. Visualization using graphics, however, is very common in universities. Especially in the field of natural sciences, complex issues are explained through various types of diagrams such as graphs, data-flow diagrams or technical drawings. Therefore, making graphics accessible to blind students is essential. Accessibility to graphics implies three things: first the blind user must know how a graphic looks like to sighted people to understand why people are discussing about visual elements of a graphic. Second, they must be able to "read" the textual or tactile representation of a graphic and to understand its meaning. Third, they must be able to produce their own graphics.

In the TeDUB project, a prototype was developed which makes technical graphics accessible [4]. The tool translated visual elements to a textual representation which can be read and produced by blind persons. Unfortunately, the system is not available anymore. There are two traditional techniques to increase the accessibility of graphics for the blind users: textual description and tactile printing. Both methods require an elaborated training for people who verbally describe or create tactile graphics. ASCII-graphics combine the two approaches. Graphical and textual elements are expressed by ASCII letters. In

electronic newsgroups, circuit diagrams are exchanged via ASCII graphics. The Study Center for the Visually Impaired Students in Karlsruhe applied this technique to make graphics accessible in the area of natural sciences for the blind students [3].

There are several advantages of drawing graphics with ASCII letters: (i) the technique is easy to learn, (ii) drawing ASCII graphics is less tedious than tactile graphics, (iii) blind students are able to read the graphics line by line on their refreshable Braille display, (iv) there is a possibility that blind students draw pictures themselves. Therefore, ASCII graphics seem to be very attractive for making graphics accessible. However, the students have to train the ability of reading such ASCII-graphics. The hardest part of reading is that one has to memorize the content of each line and mentally put the picture together.

In our approach, we aim at improving the accessibility of ASCII graphics by automatically extracting and printing them with a Braille embosser. An embossed print allows the student to get a fast overview of the complexity and the structure of a graphic. Moreover, we suggest an automatic method to disambiguate the ASCII letters by differentiating text from graphics and to print graphics in "normal font" and text in Braille as the normal fonts for graphical elements such as "—" or "/" are more similar to lines than Braille letters (dots) are (Figure 1(c)). We believe that our technique increases the attractiveness for blind students to read and draw ASCII graphics.

2 Method

In this section, we describe how we exploit hundreds of already created ASCII-graphics and how we improve their accessibility. The Study Center for Visually Impaired Students in Karlsruhe has been busy with making books and lecture notes accessible for many years. Many of the graphics are produced as ASCII-graphics. Figure 1(a) shows a circuit diagram in ASCII format [1] as sighted users see it and Figure 1(b) displays the graphic in Braille as blind users read it. However, students report problems with understanding ASCII-graphics which they read on their refreshable Braille displays. When printed on a Braille embosser students agree that ASCII-graphics can be understood more intuitively. Thus, it is desirable to provide a better access to those graphics. The automatic creation of a corpus of printable graphics increases the value of the books and lecture notes. The corpus can be used for training and evaluation in research about improving the accessibility of ASCII-graphics. Moreover, it can be used by blind students to encourage them to create their own graphics.

2.1 Creation of a Corpus of ASCII-Graphics

Recognition of ASCII-graphics in ASCII-books is simple as the drawings follow a specific syntax. The beginning of a graphic can be identified by a sequence of hyphens followed by a right angle bracket. The end of a graphic is labeled by a left angle bracket followed by a sequence of hyphens. We write a simple perl script

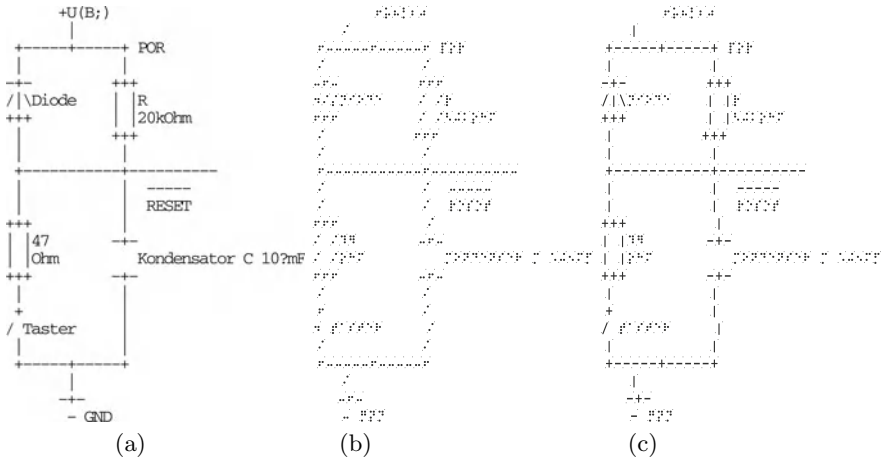


Fig. 1. Three versions of an ASCII graphic. Left: original ASCII graphic; Middle: Braille version; Right: combination of Braille and normal fonts.

that automatically filters the text files extracting the ASCII-graphics. There are different types of graphics retrieved from the data such as tables, graphs or diagrams. Additionally, we also extract the title of the files and the page number to allow an easy way of recovering the graphic while reading the digital books. We retrieved a corpus of almost 6000 graphics. Some of the graphics are described verbally or are produced tactiley. These are automatically filtered from our corpus.

3 Discussion

ASCII-graphics provide visually impaired students with a possibility of reading graphics. Printing the graphics can help blind students in different ways:

1. The prints can be used to get a fast overview of an ASCII-graphic. If the graphics are very complex the font size can be reduced. Although the script can not be read on the print, it still allows an impression of the shape and the complexity of the graphic.
2. If the prints are not too complex, the graphics can be printed in a readable size.

However, letters of ASCII graphics are ambiguous. They are either used to express text within the graphic or are used to express graphical elements. For instance, the character ">" can either describe a right arrow head or it can mean "greater than". A comparison of Figure 1(a) and 1(b) shows that the graphical part of Figure 1(a) is much closer to the original graphic representation of a figure than the Braille version. Thus, we propose to automatically disambiguate

the letters and the graphical symbols by adapting the part-of-speech tagging technique to this task. A statistical part-of-speech tagger [2] can disambiguate between these letters. We prepare a small corpus for manual annotation as basis for training and evaluating a part-of-speech tagger.

Some may argue that ASCII graphics is an outdated technique. We believe that ASCII graphics are currently a good option making graphics accessible to blind people due to their reduced learning time, platform independence, and the fact that they allow blind people to create their own graphics. Additionally, we believe that the attraction of using ASCII graphics can be increased with new tools. For instance, we suggest to use the Hyperbraille display [5] for reading and drawing ASCII graphics. The display might be used by blind students to evaluate their own graphic in the future.

4 Conclusions and Future Work

We describe a simple way of improving the accessibility of ASCII graphics by automatically extracting them from accessible ASCII-books and preparing them for the printout. We also propose to disambiguate the letters used for text and graphics respectively with a statistical tagging method. The next step is to create a manually tagged training corpus which can be used to train and evaluate a statistical tagger. Moreover, the corpus in general can be exploited to develop an automatic method to translate graphics to ASCII graphics. Furthermore, the corpus can be used to further develop sonification of technical diagrams to allow an additional access to ASCII graphics.

References

1. Bähring, H.: MICRORECHNERSYSTEME, p. 23. Springer, Berlin
2. Brants, T.: TnT - A Statistical Part-of-Speech Tagger. In: Proc. of the ANLP Conference, USA
3. Dittler, W.: Erstellung von taktilen Vorlagen nach grafischen Abbildungen am Studienzentrum für Sehgeschädigte. In: Tagungsband: Taktile Medien - Kolloquium über tastbare Abbildungen für Blinde, Freital (1995)
4. Blenkhorn, P., Evans, G., King, A., Crombie, D., Dijkstra, S., Gallagher, B., Magennis, M., Wood, J., Hagen, C., Schlieder, C., Horstmann, M., Ioannidis, G., O'Neill, A.-M., Petrie, H.: Providing Blind People with Access to Technical Diagrams. In: Proceedings of the 7th AAATE 2003 Conference, Dublin, Ireland (2003)
5. Völkel, T., Weber, G., Baumann, U.: Tactile Graphics Revised: The Novel BrailleDis 9000 Pin-Matrix Device with Multitouch Input. In: Miesenberger, K., Klaus, J., Zagler, W.L., Karshmer, A.I. (eds.) ICCHP 2008. LNCS, vol. 5105, pp. 865–872. Springer, Heidelberg (2008)

Development of a Musical Score System Using Body-Braille

Satoshi Ohtsuka¹, Nobuyuki Sasaki²,
Sadao Hasegawa³, and Tetsumi Harakawa⁴

¹ Gunma National College of Technology, 580 Toribamachi,
371-8530 Maebashi, Japan
ohtsuka@ice.gunma-ct.ac.jp

² Tsukuba University of Technology, 4-12-7, Kasuga, 305-0821 Tsukuba, Japan
nsasaki@cs.k.tsukuba-tech.ac.jp

³ Ouunkai, Body-Braille Project, 4-11-14-102, Takadanobaba,
Shinjukuku, 169-0075 Tokyo, Japan
pbb00564@nifty.ne.jp

⁴ Maebashi Institute of Technology, 460-1, Kamisadoricho, 371-0816 Maebashi, Japan
harakawa@maebashi-it.ac.jp

Abstract. We have developed a musical score system for visually impaired people and deaf-blind people using Body-Braille. The normal musical score is very convenient for non-disabled people, but it is difficult for disabled people to use it. In order to resolve this issue, 9 micro vibrators are used to express the note and the duration time of each sound, which we call “vibration score system”. This system is very suitable not only for the study of a melody but also for disabled people’s reference while playing. We performed several experiments with a subject, and in each case, we obtained successful results.

Keywords: Body-Braille, score system, vibration, visually impaired people, deaf-blind people.

1 Introduction

Since 2002, we have been developing the Body-Braille system[1], which transmits Braille characters to disabled people through vibrations on any part of the body. This year, we began to apply this technology to the field of music. Using 9 micro vibrators, any melody whose sound range is less than 2 octaves can be expressed by vibration. We call it “vibration score system”. In this system, 7 vibrators correspond to the notes (Do Re Mi Fa So La Ti) and 2 vibrators correspond to the half tone and octave. We performed several experiments with this system using the Body-Braille system. In each case, we obtained successful results for applying vibration to music expression.

2 Body-Braille

Body-Braille is a system for the presentation of a Braille character by 6 micro vibrators through any part of the body. This system is very useful for

deaf-blind people as well as visually impaired people. One of the strong merits of this system is that disabled people can keep their hands free. We have performed several experiments testing the use of B-brll (an equipment for Body-Braille) for the support of disabled people's daily life such as "Tele-support system", in which sighted people support disabled people through the TV telephone system of a G3 mobile phone, "Independent walking support system" using IC-tags embedded in roads to supply information to disabled people during transportation, and "Phone system for deaf-blind people", which enables real-time conversation between deaf-blind people through B-brll and DTMF tone signals in just the same way as the audio conversation between ordinary people. In this paper, we explain how we have applied B-brll technology to the entertainment field.

3 Equipment

The size of the B-brll equipment is 160mm(W), 40mm(H), 100mm(D) and the weight is 500g. It can drive 12 micro vibrators corresponding to 2 braille characters. Figure 1 shows the exterior view of the equipment. We use the PIC chip from Microchip Corporation as a CPU. Software is developed for each application and an eprom in the CPU chip is rewritten by a rom writer each time.



Fig. 1. Exterior view of Body-Braille control device

4 Music Application

Although a Braille score system has already been developed, it is not easy to master and it is not possible to refer to the score while playing a musical song. So we began to study how to use B-brll technology for the presentation of a music score. If the music score can be presented by only vibration, it would be useful for not only deaf-blind people but also visually impaired people. And it can be used for not only studying the music melody but also for reference while playing because in the B-brll system, users don't have to use their hands to reference the score.

5 Implementation of the Vibration Score

To implement the vibration score system, we use nine out of twelve micro vibrators. Seven micro vibrators are used for the notes (Do Re Mi Fa So La Ti), one vibrator is used for the half tone and the other vibrator is used for octave presentation. Table 1 shows the assignment of 9 vibrators. The length of sound is expressed by the vibration continuation time. If the sound range of a melody is less than 2 octaves, any melody can be expressed by vibration. We call it the “vibration score system”. The half tone and octave vibration occur simultaneously with the note vibration, but to detect plural vibration easier, the duration of these two vibrations is set to half of the time of each note. To produce the vibration score, we use the free MIDI software, “Muse”, to input the normal music score using only text information. Then using special software developed by us, the Muse score data is converted to the vibration score which uses two bytes for one sound (one for note, the other for continuation time). Figure 2 shows the detailed configuration of the vibration score. Finally this vibration score data is written in the eeprom of the CPU.

Table 1. The Assignment of 9 vibrators

Do	Re	Mi	Fa	So	La	Ti	half tone(#)	octave
1-①	1-②	1-③	1-④	1-⑤	1-⑥	2-①	2-②	2-③
Little	Ring	Middle	Index	Thumb	Thumb	Index	Middle	Ring
left hand					right hand			

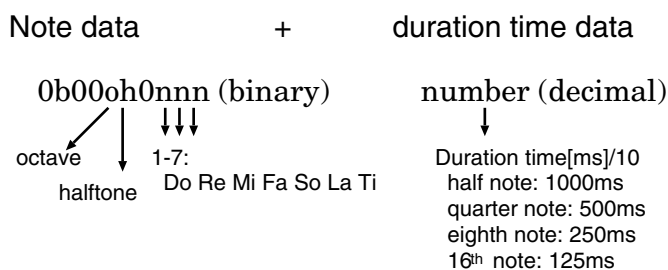


Fig. 2. The Detailed Configuration of the Vibration Score

6 Experiment

The subject was a blind girl whose hobby is playing and composing music. After being presented with the vibration of a part of a melody via B-brll, she played an electronic piano to check the melody. Easy Japanese nursery songs which are



Fig. 3. Experiment-1



Fig. 4. Experiment-2

very famous and common knowledge were used. Out of a total of 61 songs, 31 songs were recognised after less than two presentations, 15 songs were recognised after less than four presentations and 15 songs were recognised after less than ten presentations. In a second experiment, classic Japanese folk songs which are unfamiliar to most young people were used. As for 5–6 songs, the subject could master and play the complete melodies in less than 60 minutes study through the vibration score. Figures 3 and 4 show how the experiments were done.

7 Conclusion

We tried to apply the B-brll technology to the entertainment field, especially in the presentation of music scores. Using 9 micro vibrators, any melody whose sound range is less than 2 octaves can be expressed by only vibration. Through several experiments, we could obtain successful results for applying vibration to music expression. In the near future, we would like to conduct an experiment of the dynamic use of the vibration score system, that is, referencing the vibration score while playing.

Acknowledgment

A part of this work was supported by JSPS KAKENHI 20500497.

Reference

1. Ohtsuka, S., Sasaki, N., Hasegawa, S., Harakawa, T.: Body-Braille System for Disabled People. In: Miesenberger, K., Klaus, J., Zagler, W.L., Karshmer, A.I. (eds.) ICCHP 2008. LNCS, vol. 5105, pp. 682–685. Springer, Heidelberg (2008)

Analysis of Awareness while Touching the Tactile Figures Using Near-Infrared Spectroscopy

Takayuki Shiose¹, Yasuhiro Kagiya¹, Kentaro Toda¹, Hiroshi Kawakami², Kiyohide Ito³, Akira Takatsuki⁴, and Akitoshi Seiyama⁴

¹ Kyoto University, The Kyoto University Museum

² Kyoto University, Graduate School of Informatics

³ Future University-Hakodate

⁴ Kyoto University, Graduate School of Medicine

shiose@inet.museum.kyoto-u.ac.jp,

{kagiya, kawakami}@sys.i.kyoto-u.ac.jp,

itokiyo@fun.ac.jp, {atakatsu, aseiyama}@hs.med.kyoto-u.ac.jp

Abstract. In this paper, we propose a method to measuring how the blind to touch the tactile images using Near-Infrared Spectroscopy (NIRS). The tactile image is one of powerful tool to give any visual information for the blind. On the other hand, it is difficult to estimate the process that the blind has understood the tactile images because the only trajectories of the hand are too vague to identify when the blind know the target image. Some experimental results insist that the brain is strongly activated at beginning of a task rather than the process of the task. Then, NIRS is confirmed to be useful to analyze process of understanding tactile figures by touching.

Keywords: assessment, design for all, blind people, neural science.

1 Introduction

The purpose of this study is to investigate how much visually impaired people actually understand while they are touching tactile graphics. Recently, we reported about Inclusive Design Workshops [1][2]. In workshops based on Inclusive Design Method, visually impaired and elderly people play active role in design processes. Therefore, support of communication between participants is important. For this purpose, we developed the system which helps visually impaired people to understand landscape photographs and micrographs [2]. First, it processes photographs to simplify them into line images, and then, make them tactile by a stereo copying machine.

To investigate how much visually impaired people understand tactile figures, we analyzed movements of their hand on tactile figures [2]. As a consequence, it was found that frequency of large movements of a hand relates complexity of a tactile figure. In this work, we measured brain activity more precisely at similar task (touching tactile figures) by NIRS (near-infrared spectroscopy) brain measuring system, which is one of non invasive functional mapping systems with multi channels [6]. By using NIRS, amount of Hb (hemoglobin) at each areas of a brain can be measured by measuring reflection of NIR (near infrared) from the scalp.

2 Experiment

There were 3 visually impaired subjects and 3 sighted subjects (between 20 and 50 years old). We measured activity of prefrontal area of all subjects when they were touching tactile figures. In the case of 2 subjects, activity of occipital area was also measured.



Fig. 1. A subject with NIRS. IR receptors were placed on subject's scalp and their positions correspond to cortical areas.

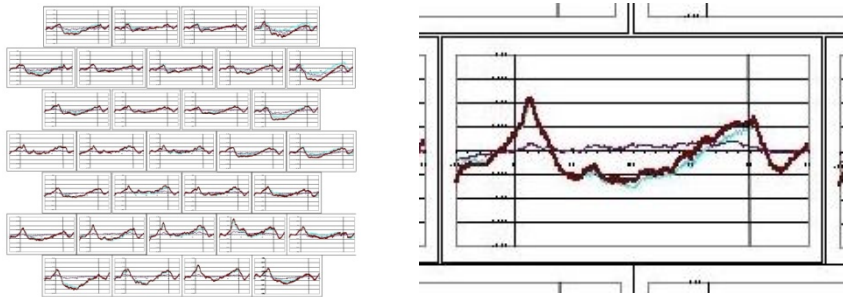


Fig. 2. Variations of Hb concentration at each channel. The right figure represents the enlargement of the graph of the 24-th channel.

Subjects were requested to guess what is drawn on a tactile figure by only touching it. On each tactile figure, an alphabet, a hiragana (Japanese character), or a sign was drawn. 9 tactile figures on A4 paper were shown to each subject. Subjects touched them in 60 seconds and rested 15 seconds before and after each touching.

The left of Fig. 2 shows variations of Hb concentration in a task at each channel, averaged in all subjects. The bottom row corresponds to the front of heads. In this

figure, a variation at channel 24 (the right of Fig. 2) is distinctive. At channel 24, Hb concentration increased immediately after beginning of a task and then decrease to lower than value in resting phase. This result differs from general results of brain measurements, which shows that a brain is strongly activated at beginning of a task [4][5].

Moreover, the result of a subject is notably different from others' at channel 29. Fig. 3 shows results when hiragana "wa", "a", and "ma" was shown consecutively to him, from the left to the right.

In the first case ("wa"), he might be touching without understanding what was on the tactile figure, however, the value increased at latter part. In contrast, in the second case ("a"), this area of his brain was strongly activated. We can conjecture that it was activated when he recognized that what he was touching was a letter, which has a meaning, was not meaningless figure.

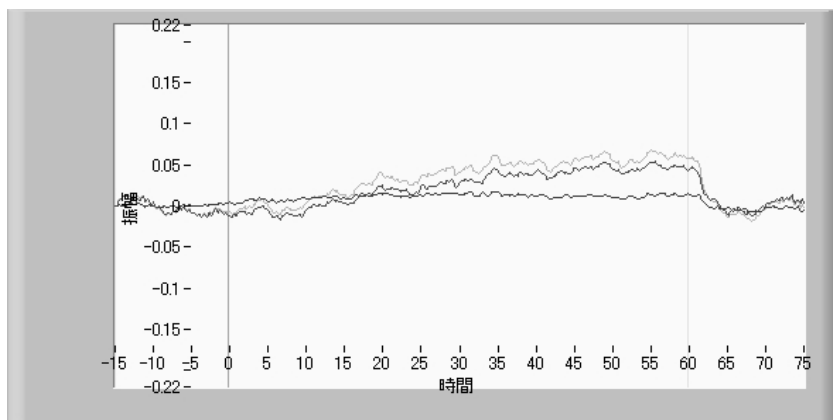


Fig. 3. The process of the moment when a subject comes to know what he touched

3 Discussions

Several researches show that some areas of a brain are deactivated while one doing a task. Raichle focused on brain activities in resting phase and called them Default Mode Network [3]. The result at channel 24 is similar to these phenomena; it is possible that a part corresponding channel 24 works as Default Mode Network in the task to guess by only touching, not by seeing.

Although Default Mode Network is attracting attention in brain science, it is not fully understood. Hence, we must investigate further to judge whether each area of brain is activated or deactivated when a subject start or end a task. Nevertheless, it seemed reasonable to think that difference of Hb concentration, not value itself, reflects well change of recognition. Therefore, NIRS is useful to analyze process of understanding tactile figures by touching.

References

1. Shiose, T., Toda, K., Kawakami, H., Katai, O.: Inclusive Design Workshop by Regional Cooperation between an NPO and a University. In: Ishida, T., R. Fussell, S., T. J. M. Vossen, P. (eds.) IWIC 2007. LNCS, vol. 4568, pp. 355–367. Springer, Heidelberg (2007)
2. Shiose, T., Kagiya, Y., Ito, K., Kawakami, H., Katai, O.: Toward touching a Landscape in a Picture: Investigation of Grouping Strategy about tactile Images and Image Simplification Method. In: Miesenberger, K., Klaus, J., Zagler, W.L., Karshmer, A.I. (eds.) ICCHP 2008. LNCS, vol. 5105, pp. 859–864. Springer, Heidelberg (2008)
3. Raichle, M.E., MacLeod, A.M., Snyder, A.Z., Powers, W.J., Gusnard, D.A., Shulman, G.L.: A default mode of brain function. *PNAS* 98(2), 676–682 (2001)
4. Vega-Bermudez, F., Johnson, K.O., Hsiao, S.S.: Human Tactile Pattern Recognition: Active Versus Passive Touch, Velocity Effects, and Patterns of Confusion. *Journal of Neurophysiology* 65(3), 531–546 (1991)
5. Noguchi, Y., Takeuchi, T., Sakai, K.L.: Lateralized Activation in the Inferior Frontal Cortex during Syntactic Processing: Event-Related Optical Topography Study. *Human Brain Mapping* 17, 89–99 (2002)
6. Schroeter, M.L., Zysset, S., Kupka, T., Kruggel, F., Yves von Cramon, D.: Near-Infrared Spectroscopy Can Detect Brain Activity During a Color-Word Matching Stroop Task in an Event-Related Design. *Human Brain Mapping* 17, 61–71 (2002)

Development of Directly Manipulable Tactile Graphic System with Audio Support Function

Shigenobu Shimada¹, Haruka Murase², Suguru Yamamoto, Yusuke Uchida,
Makoto Shimojo², and Yutaka Shimizu

¹ Tokyo Metropolitan Industrial Technology Research Institute, 3-13-10 Nishigaoka,
1158586 Tokyo, Japan

² The University of Electro-Communications, Department of Mechanical Engineering and
Intelligent Systems, 1-5-1 Chofugaoka, 1828585 Tokyo, Japan

Abstract. A basic device combining a tactile graphic display function and a touch position sensing function is proposed. The trial device consists of two major components, a tactile graphic display and a force sensor. The force sensor measures a center of gravity generated by touch action on the display surface and a PC estimates the point based on the data. The fundamental function for achieving the interactive communication is to detect the location where user is touching a tangible surface. By applying this functions, the click and scroll function by an empty hand are realized. In addition, an audio-tactile graphic system which can be used mainly to overcome tactile cognitive limitation is implemented as an application of the system. Moreover, the area of tactile display is expanded for the purpose of improvement usability. Following this, we propose the tactile scrollbar where the users can recognize their position of plane space. The validity of the developed tactile graphic system has been confirmed through subjective experiments.

1 Introduction

Most of the tactile graphic displays developed in the past were simply unidirectional communication tools [1,2,3]. If a bidirectional communication function could be included in these displays, it would not only lead to graphic information becoming more accessible to visually handicapped persons but would also stimulate their creativity. Keeping this in mind, we developed an interactive tactile graphic display that can be directly manipulated by hand [4]. The device consists of two major components, a tactile graphic display and a 6-axis force torque sensor. The 6-axis force torque sensor measures the force which the user applies to the tactile graphic display, and personal computer (PC) estimates the touch position on the tactile graphic display which is applying force based on the measurement value. Since the abovementioned interactive tactile graphic display can be used to provide spatial information to visually handicapped people, it can serve as an alternative to a simple tactile graphic display. In a previous study, Gardner et al. developed an interactive braille figure using a braille figure printer, a PC, and a digitizer (tablets) [5,6]. If a user touches the braille figure with variable gradation

and a fine pitch, the device detects the touch position from the locus of user's finger. The user can thus obtain information about the braille figure and its accompanying audio. However, since a hard copy of the braille figure is used, there are some problems associated with this system: printing the braille figures requires extensive time, the printing process is expensive, and finally, the figures cannot be zoomed in and out. Therefore, in order to incorporate audio feedback in our apparatus, we developed a tactile graphic system with an interactive function [7]. In the experimental study using above device, a good subjective assessment was obtained but some points could be improved. First, the size of this device should be as large as possible. Visually handicapped persons have been educated to touch with both hands for the tactile reading. Therefore, the operability of this device is poor because the size, 32 × 48 pins, is a little small for people to touch in that way. Second, when part of the entire figure displayed on the display of a PC is projected on the tactile display, the relationship between the "entire figure" and "the part" is unclear. The projected area of the tactile display is much smaller than the display area of a PC. Therefore, when a figure is larger than the projection area, users must use the arrow key and the scroll function of the tactile graphic system for the touch reading. In this case, they carry out the touch reading to the Braille figure imagining the movement of the projection area. However, that work is very difficult and users often lose their position and have to restart from the origin. In addition, the pins on the tactile display are only raised or withdrawn. So, when areas with nothing displayed or when the area is the same color, there are no clues for touch reading.

Consequently, this paper describes the development of a 32 × 96 tactile display, doubled from the conventional 32 × 48 display. Because the use of a 6-axis force torque sensor deteriorates the accuracy of estimating the positions of fingers, we devise and design a force sensor appropriate for the display. We demonstrate the new system with the same finger position estimation capability as the previous system on a trial basis. Moreover, We propose a new tactile scrollbar using the motorized slider to show the position of projected area which is part of the display of a PC. We conduct a basic experiment to test whether the position could be determined from the tactile scrollbar and executed a comparison experiment to compare the changes in position identification accuracy with and without the tactile scrollbar in order to confirm the utility of the newly developed system.

2 Target System

2.1 Previous Directly Manipulable Tactile Display

The previous tactile graphic display consists of 4 bimorph modules (SC.10, provided by KGS Inc.). Each module has a 32 × 12 arrangement of pins with a spacing of 2.4 mm. The total tactile surface has 32 × 48 arrangements activated at a particular time. The pin is 1.3 mm in diameter, and the height of the pin when activated by a data signal is 0.7 mm. The refresh time for the entire surface is approximately 50 ms. The device interfaces with the host PC through USB communication. A 6-axis force torque sensor is used for the previous system. The force and the torque generated by touch action is measured by the sensor. Based on the measured data, the touch position of user is estimated. The sensor is set roughly below the center of the display component. Newly

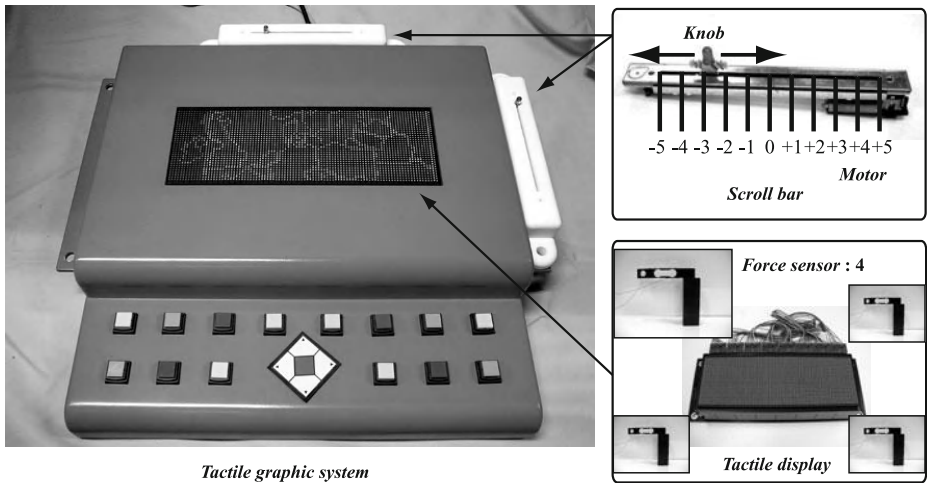


Fig. 1. New developed tactile graphic system

Table 1. Comparison between previous and newly developed device

device	previous	newly developed
pins	1536	3072
position estimation sensor	6-axis force/torque	one-axis force
show a global position	no use	use (tactile scroll bar)

developed tactile graphic system comprises eight modules (see the left **Fig.1**). The area of the tactile display is 32 × 96 pins. Therefore, we prepared a one-axis force sensor shown in the right lower **Fig. 1** to replace the 6-axis force torque sensor for improved estimation accuracy, lower price, and lighter weight. The comparison between previous and newly developed device is shown in **Table 1**.

2.2 Application for Tactile Map with Audio Function

The core of the tactile map system is a program made in Flash of Adobe Co. (the following Flash program). The main role of the Flash program is to convert the contents file into each type of visual display as coloring image and tactile display as binary image. The contents file has colored image of the every region and correspondence table between color and audio file. The Flash program executes the audio file in proportion to the color which the click coordinate allocated, when it clicks the region with the tactile display. Color itself does not have the meaning. The conceptual scheme is shown in **Fig. 2**. The file of image and audio used by this system is loaded in the execution with the arrangement to the outside without embedding for the Flash program. Therefore, there is no necessity of amending the Flash program in the new contents preparation. The use file type is general, as the image uses the PNG, and as the audio uses MP3. So,

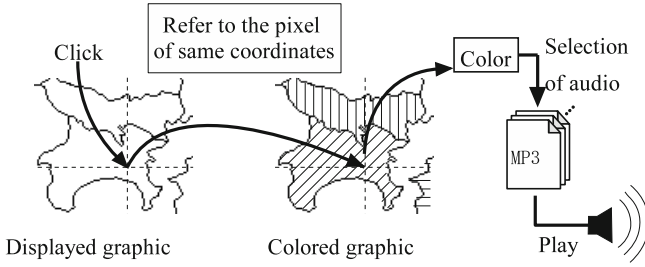


Fig. 2. The schematic drawing of location/audio mapping method

exclusive software and procurement difficult software are not required for preparation and editing of contents. This is one of the feature of this system.

3 Development of Force Sensor

The right lower **Fig. 1**. shows the force sensor we manufactured. The sensor, made of duralumin (A7075), was designed to detect a full 25 N and consists of a cantilever beam with a parallel plate structure and a strain gauge to detect the bend stress caused by a load. We arranged this sensor in each corner of a structure holding the tactile display. We prepared the tactile graphic system using four corrected force sensors. We placed 420 gf weights on 13 points in ascending order determined on the tactile display and recorded each point output from the processing unit as an estimation point (see the left **Fig. 3**). We recorded a set of these 13 points five times in total. Because the pitch between pins (i.e., resolution) is 2.4 mm, tolerance of up to 2.4 mm was permitted between the loaded point and the estimation point. The right **Fig. 3**. shows the plots between these points. The pitches between these points were accurate within 1.3 mm on average and 3.5 mm at maximum. The result showed that the newly developed system met the design requirement.

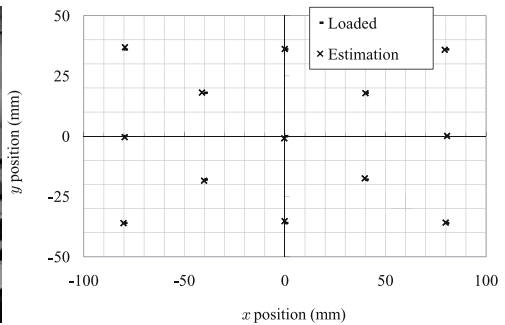
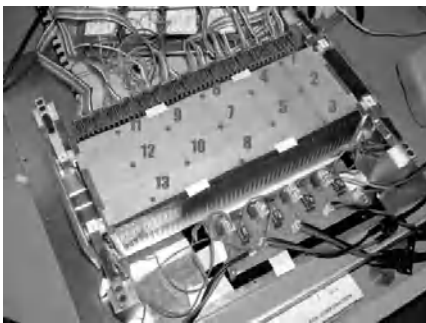


Fig. 3. Results of experiment : the left and right figure shows the measurement points and the experimental result, respectively

4 Development of the Tactile Scrollbar

The scrollbar used with a PC is a GUI component. It informs the user which data are displayed when part of the data extends beyond the displayed window in a window system. The position of the part of the entire data is shown by the position of the knob inside the scrollbar. Similarly, in our basic concept, the above displayed window in a window system is the projected area of the tactile display, and the above knob is the knob of tactile scrollbar as shown in **Fig. 4**. For this purpose, we developed a tactile scrollbar using a motor-driven slide volume shown in the right upper **Fig. 1**. This is a rectilinear-movement volume and the knob is driven by the motor. Since part of the figure on the display of a PC is projected to the tactile display as an embossed pattern, the position of the knob of the tactile scrollbar corresponds to the absolute coordinate of the centroid of the projection area on the display of a PC. As shown in the left **Fig. 1**, it was equipped with the tactile display vertically and horizontally.

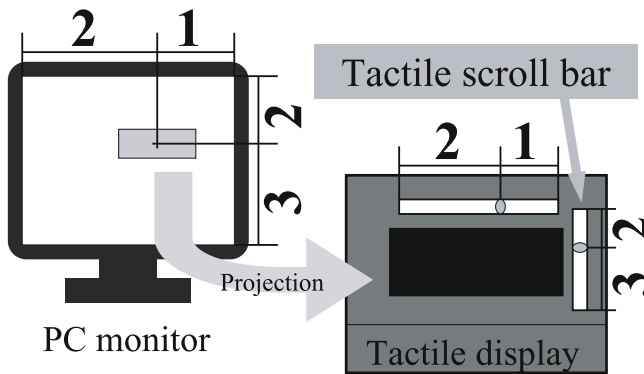


Fig. 4. The schematic drawing of tactile scrollbar

4.1 Experiment for One-Dimensional Position Discrimination

We conducted a basic experiment using only one vertical horizontal axis to show how a user could judge 11 points by touching the knob of the tactile scrollbar. We also compared accuracy with and without the resin domical indicator (diameter : 2.5 mm; height : 1.5 mm) in the middle of the scrollbar to improve accuracy. We determined 11 points (from -5 to 5; the middle point is 0) obtained by dividing the scrollbar by 10, assuming the knob stopped at each point. We moved the knob to any of these points randomly and asked blindfolded test subjects to touch it and indicate the position of the knob (11 points from -5 to 5) and the level of confidence in their choice (5 degrees from 1 to 5). We recorded the time required for them to answer. We conducted two sets of 33 trials vertically and horizontally. We selected six sighted students as test subjects. **Fig. 5**, **Fig. 7** show the result of the test above. The left and right sides of each figure show the results with and without the indicator. The horizontal axis of all figures shows the position of the knob. **Fig. 5** shows the correct answers rates. The system turned out

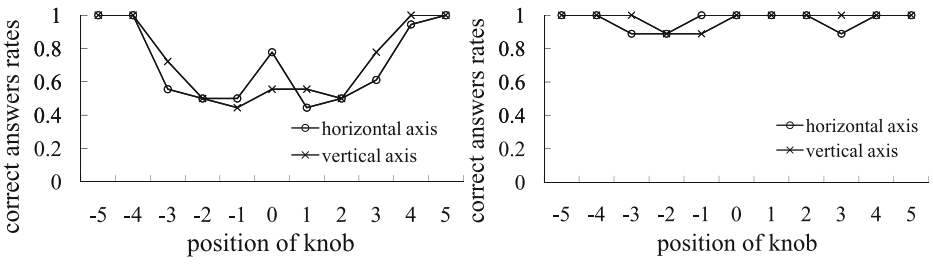


Fig. 5. Correct answers rates : the left and right figure shows with and without indicator, respectively

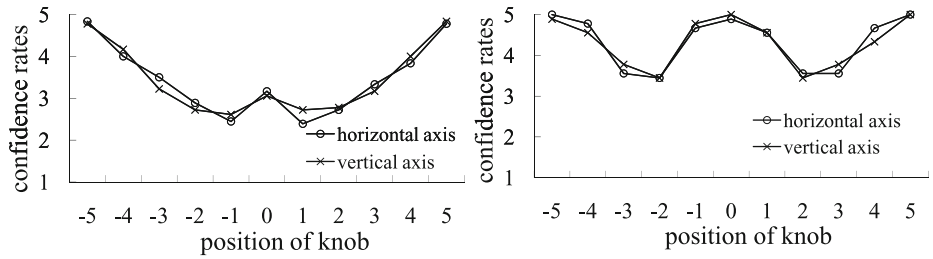


Fig. 6. Confidence rates : the left and right figure shows with and without indicator, respectively

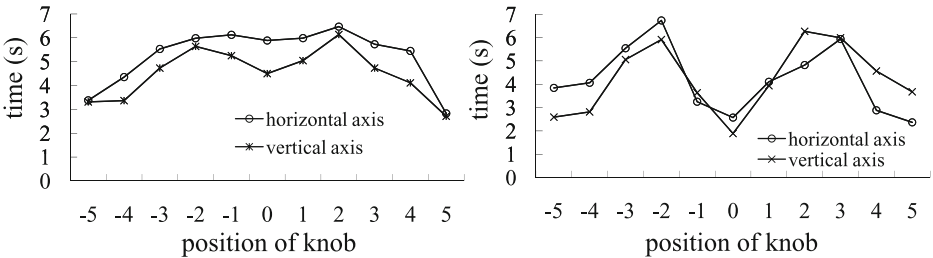


Fig. 7. Task completion time : the left and right figure shows with and without indicator, respectively

to be sufficiently feasible with the indicator. We can see the effectiveness of the indicator when comparing both figures. Fig. 6. shows the level of confidence of the test subjects. The effectiveness of the system is remarkable when the indicator is used, and the level of confidence is three or higher according to the correct answers rates. Fig. 7. shows the task completion time. The task completion time is considerably shorter around the middle point with the indicator while it is longer at 3 2 2 and 3. Considering the results without the indicator (left side of each figure), the 0 position in the horizontal direction is higher compared to the vertical direction at the correct answers rates, but it is not reflected in the level of confidence. Because actions are different in the vertical and horizontal directions, we can see that the test subjects felt differently depending on the direction. After seeing the results with the indicator (right side of each figure), simply by adding one indicator improved the correct answers rates, level of confidence, and

task completion time in the experiment. We found that one method for obtaining higher recognition accuracy was to add one indicator, but the trade-off was task completion time.

4.2 Experiment for Two-Dimensional Position Discrimination

The use of the tactile scrollbar enabled users to determine a one-dimensional position. Then, we studied whether plane positions could be recognized by using the scrollbar both vertically and horizontally. **Fig. 8.** is a conceptual diagram of the experiment. We prepared an figure area of 256 x 192 pixels where we arranged an object (circle) and five dummies (triangles) randomly. The figure area was so wide that it could not be projected on the tactile display at one time. We asked blindfolded test subjects to search the figure by moving the arrow key over part of the projected area. When the test subjects found the object by the touch reading on the tactile display, they recognized their position on the entire figure area using the tactile scrollbar. The test subjects made marks on the answer sheet above their recognized position, then we asked the subjective usability of the scrollbar (5 degrees from 2 to 2). We assessed the scrollbar based on errors in position between the object and the markers and their search time. We prepared three figures for the experiment and studied cases with and without the scrollbar for each figure. We selected 10 test subjects. **Table 2.** shows the average errors and search time with and without the tactile scrollbar. The t-test revealed that both the average errors and search time were significantly different at a level of 5%. Most of the test subjects answered, "the system can be used more easily with the tactile scroll bar," and the average degree of the subjective assessment was 1.87.

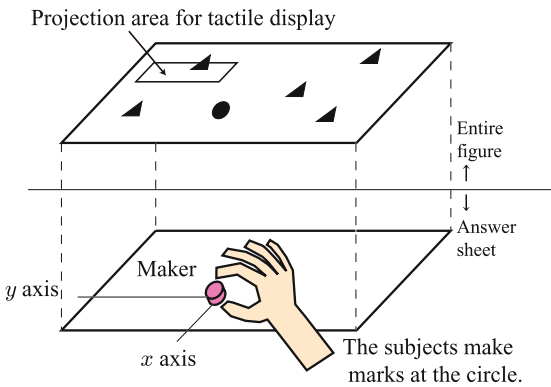


Table 2. Relationship between average errors and search time

scrollbar	average errors (mm)	search time (s)
use	62.8	142.8
no use	79.3	199.6

Fig. 8. Answer sheet for experiment of two-dimensional position discrimination

5 Conclusion

Through this study, we found issues with the previous findings and developed hardware suitable for a new target system. The newly developed force sensor detected one-axis

force, achieving the same finger position estimation accuracy as the previous system. The simple structure is expected to achieve a lower price and lighter weight. In addition, the double size display did not decrease estimation accuracy. We can expect the sensor to be used in larger displays and applied to consumer touch panels. The issue of losing touch reading points in the application was solved by developing a tactile scrollbar that showed the global position. Because the scrollbar moves the projected area on the tactile display by actively moving the knob to the desired position and shows the global position, the interface is highly usable. We successfully improved usability with the global position detection function and the larger display in addition to the scroll, click, and scaling functions for the direct manipulable tactile graphic system. In light of these functions and supporting network, we will develop applications for education, such as graphs and text layout presentations, in addition to the already developed tactile map applications, as tactile GUI devices for visually handicapped persons.

References

1. Wellman, P.S., Peine, W.J., Favalora, G., Howe, R.D.: Mechanical design and control of a high-bandwidth shape memory alloy tactile display. In: Bibel, W., Jorrand, P. (eds.) *Fundamentals of Artificial Intelligence*. LNCS, vol. 232, pp. 56–66. Springer, Heidelberg (1986)
2. Shinohara, M., Shimizu, Y., Mochizuki, A.: Three-dimensional tactile display for the blind. *IEEE Trans. on Rehabilitation Engineering* 6(3), 249–256 (1998)
3. Wagner, C.R., Leederman, S.J., Howe, R.D.: A tactile shape display using RC servomotors. In: *Proc. of Satellite Meeting of the Annual IEEE VR 2002*, pp. 354–356 (2002)
4. Shimojo, M., Shinohara, M., Tanii, M., Shimizu, Y.: An approach for direct manipulation by tactile modality for blind computer users: Principle and practice of detecting information generated by touch action. In: Miesenberger, K., Klaus, J., Zagler, W.L., Burger, D. (eds.) *ICCHP 2004*. LNCS, vol. 3118, pp. 753–760. Springer, Heidelberg (2004)
5. Gardner, J., Bulatov, V.: Complete Access to All Paper and Computer Forms, Tables, and Charts. In: *Proc. of the 2003 CSUN International Conference on Technology and Persons with Disabilities*, pp. 17–20 (2003)
6. Gardner, J., Bulatov, V.: Scientific Diagrams Made Easy with IVEO. In: Miesenberger, K., Klaus, J., Zagler, W.L., Karshmer, A.I. (eds.) *ICCHP 2006*. LNCS, vol. 4061, pp. 1243–1250. Springer, Heidelberg (2006)
7. Shimada, S., Yamamoto, Y., Shimojo, M.: Scientific Diagrams Made Easy with IVEO. In: *18th IEEE International Conference robot human interaction (RO-MAN 2009)*, pp. 1243–1250. IEEE, Los Alamitos (2009)

AutOMathic Blocks: Supporting Learning Games for Young Blind Students

Arthur I. Karshmer¹ and Ken Paap²

¹ University of San Francisco
San Francisco, CA USA
akarshmer@usfca.edu

² San Francisco State University
San Francisco, CA USA
kenp@sfsu.edu

Abstract. The AutOMathic Blocks system has been developed to help young blind students in the acquisition of elementary math skills. Through the use of Braille labeled blocks, a plastic grid, a touchpad device and any computer, the system has been defined to aid the young student in this most important domain of education. Without these basic skills, students have a much higher probability of not being able to enter math related professions.

Keywords: Mathematics, Learning, Blind.

1 Introduction

In general, the acquisition of math skills is dramatically more difficult than for sighted students. The fact that even introductory arithmetic is a two dimensional representation, complicates the learning ability of young blind students. More advanced math is often what is called a $2\frac{1}{2}$ dimensional representation – even more difficult for the blind. Even Braille representations of math present the subject in a linear fashion.

The AutOMathic Blocks [1] system obviates these problems by presenting early math in its true form of non-linearity. Through the use of blocks labeled with simple Braille, the learner can tactically examine math problems with their fingers and thereby receive a two dimensional view of their work. It has been demonstrated that tactile input can do the equivalent of mapping tactile information to the visual cortex.

Learning does not end with formal representations of information. Over the recent past, games have been implemented to aid in the learning process. To that end, we have begun to implement simple game playing into the AutOMathic Blocks system. The first game implemented is the traditional tic-tac-toe game for two players. It will be used to help blind students learn how to work in an environment that resembles the environment used by their sighted colleagues. As a further benefit, it will enable cognitive scientists to learn more concerning object representation in the mind of the congenitally blind. The remainder of this paper describes several experiments using the games built into the AutOMathic Blocks system to evaluate the usefulness and efficacy of the presentation of math in a two-dimensional form to help visually

impaired children learn the critical basics that will prepare them for more complex math later in the educational process.

2 Testing the Conceptual Model Using

2.1 Why Is It Particularly Difficult to Learn Mathematics without Vision?

Although visually impaired children are developmentally delayed in a variety of knowledge domains the delay is greater for those skills that require the acquisition and maintenance of a two- or three-dimensional problem space in working memory (e.g., the visual-spatial sketchpad component, [2]). This follows because vision directly provides a spatial representation where different identifiable objects can be simultaneously and precisely located in specific locations. Visually perceived space is also very flexible with respect to the way in which objects are attended to and scanned. In contrast, audition (and especially speech) relies heavily on sequential processing. Because even basic mathematics is multidimensional it is very difficult to understand and solve problems with input-output mediums that are inherently linear. Unfortunately computer-based assistive technologies have relied on translating sequences of visual objects (i.e., numbers, letters, symbols, words) to speech. The conceptual model underlying the design of our basic studies and of our prototypes assumes that touch or some combination of touch and audition is the best substitute for providing a multidimensional spatial representation in the absence of vision.

2.2 Spatial Representations in Vision and Touch

Studies aimed at new enabling technology for those with severe visual impairment is always constrained by the number of available participants, especially those who are fluent in Braille. Some of the empirical work we propose uses blindfolded participants with, otherwise, normal vision. This decision goes beyond expediency, as we are interested in basic questions regarding the ability of touch and vision to support mental representations of two-dimensional space. This requires an understanding of the relationship between touch and vision and how that relationship develops. As argued below, this knowledge has implications for differences between congenitally blind populations versus those advantageously blind and why their optimal assistive technology may differ. This section briefly summarizes what is known about the relationship between touch and vision.

A mistaken reason to believe that touch is the best substitute for vision is the common belief that visual perception is learned by concomitant experience with touch during infancy. Influential proponents of this view start with Berkeley [3] and Dewey [4] and remain accepted truths by many optometrists and ophthalmologists. However, classic experiments by Rock [5] and his colleagues have shown that when vision and touch are placed in conflict vision captures and dominates touch. With appropriate optical devices inducing a conflict between vision and touch, these experiments show that true size can be underestimated by half, that straight rods can feel curved, and that a square can feel like a narrow rectangle.

A better basis for the expectation that touch leads to mental representations of space similar to those constructed through vision is that some spatial illusions occur in

both modalities especially when the experience is produced by active haptic sensing rather than passive tactile sensing (viz., pressing a raised figure onto the skin). Not only does the Muller-Lyer illusion occur when participants freely move their fingers and can simultaneously touch the figures in several places at the same time Casla [6], Millar [7] and Wong [8], but the strength of the illusion varies in the same way as a function of the angle of the arrowhead with respect to the shaft Over [9]. These experiments with sighted, but blindfolded individuals must be interpreted carefully in light of the work described in the previous paragraph showing the vision dominates touch. That is, it may be the case that sighted individuals construct spatial representation from touch by “visualizing” in ways that congenitally blind individuals could not. It has been reported Bean [10] and Hatwell [11] that the congenitally blind also experience the Muller-Lyer illusion, but it would be informative to compare blindfolded sighted and/or advantageously blind individuals to congenitally blind participants on tasks that require the construction of spatial representations that are more similar to those of interest to this project. Another line of studies that supports the similarity of spatial relationships constructed from vision and touch is inspired by Kosslyn’s [12] seminal work on visual imaging. All the major experimental effects that supposedly reveal the spatial and non-verbal properties of visual imagery (such as mental rotation, scanning, size/inspection time effects, and selective interference), have now been demonstrated in congenitally blind participants Marmor [13], Carpenter [14], Kerr, [15], Hampson [16], Vecchi [17, Barolo [18] and Aleman [19].

Roberta Klatzky and her associates (e.g., Klatzky [20] have developed a sophisticated model of how both egocentric and allocentric representations of manipulatory space (within arm’s reach) can be generated from kinesthetic inputs. An index finger can be returned to a previously sensed target location with impressive precision. A more complex representation that computes second-order parameters allows participants to reproduce (at a new location) either the straight-line distance between two previously and simultaneously sensed finger locations (e.g., left and right index fingers) or the angle of that straight line with respect to the participant’s frontal plane with somewhat less precision. Most of the work on manipulatory space involves passive tactile or kinesthetic sensing rather than the active movements that occur when users manipulate the AutOMathic Blocks. Spatial representations constructed from active haptic sensing and prehension should be more precise. Another important difference between this body of research and our concerns is that previous studies on manipulatory space is surprisingly bereft of objects to manipulate and integrate into the spatial map. This is not true of the large body of work on locomotor space that is concerned with way finding and cognitive maps.

One series of experiments Gaunet [21] and Thinus [22] have investigated the spatial relations of haptically identified objects in manipulatory space. For example, objects with cross-sections corresponding to squares, circle, and triangles were spontaneously explored in a first phase. Following a spatial change participants reexplored the layout and had to verbally describe the modification. Although early blind participants made more errors than both late blind and blindfolded sighted groups the difference in performance appears to be caused by differences in exploratory patterns. Early blind participants relied more on visiting a sequence of multiple objects whereas the other groups relied more on moving back and forth between the same two objects several times before moving to another object. The authors speculate that the coexistence of

these two components (large exploratory cycles and back-and-forth movements) may be interpreted as corresponding to a strategy that consists of roughly comprehending the holistic configuration and subsequently refining this knowledge through more detailed analysis. Because early blind participants used a different mix of strategies the differences in performance may result more from differences in exploratory patterns than any immutable limitation on their mental representations of objects in two-dimensional space.

3 Further Support for the Conceptual Model

In summary of the previous section, the research on active haptic sensing is consistent with the hypothesis that blind or blindfolded participants can construct a good mental representation of the layout of objects in two-dimensional space, but the ideal basic research remains to be done. The ideal would map onto the environment of the AutOMathic Blocks system and consequently include these features:

- a. Employ manipulatory space rather than locomotor space
- b. Identification of objects (attributes, shapes, or characters) that fit on a fingertip
- c. Spatial relationship between objects is task relevant
- d. Spatial representations are maintained in working memory, but can be refreshed or updated through haptic exploration

These critical features are present in blindfolded versions of the familiar tic-tac-toe game that we propose to use as an experimental paradigm and is operational on the AutOMathic Blocks system. The basic question is the relative ability to form a mental representation of two-dimensional space either based on touch (Touch Alone condition) or a combination of touch and voice (Touch+Voice condition) when visual information is not available. A third condition (Abstract Coordinate) uses speech to specify coordinates and requires a more abstract (amodal) spatial representation that is not directly supported by sensory input from physical locations.

In the Touch Alone condition the participant places one of his/her raised “X” tiles in one location of a 3x3 grid that is similar to the larger workspace grid of the AutOMathic Blocks prototype. The “opponent” (experimenter) responds by placing a raised “O” tile in an open location and then passively moving the participant’s finger to that location. To keep the game interesting and the participant motivated, the opponent’s responses are based on a script that controls for the probability of the participant winning. The participant is free to review the position of the tiles on the board at any time when it is his or her turn. Dependent measures include average decision time (turn duration), the number of times the identity or location of previously played pieces are reviewed, and outcome (proportion of wins, ties, and losses). The Touch Alone condition is, of course, a scaled down version of the AutOMathic Blocks Touch Alone prototype. Having only two characters (X’s and O’s) enables one to bypass the complications of Braille fluency, to use readily available sighted participants, and to use a less complex task (compared to math problems) that, nonetheless, places a premium on maintaining a two-dimensional spatial representation of the location of several objects.

In the Touch+Voice condition the raised tiles with X's or O's are not used. Rather, each location on the 3x3 grid contains a soft response key. When the participant presses a key in a specific location s/he says "X" and a voice recognition system records an X at the pressed location. On the opponents turn the participant's finger is passively guided to the response location and the opponent says "O" recording the identity of the character at that location. On his or her turn the participant can review the contents at any location by pressing the key and hearing "X", "O", or "blank". Note that this condition involves the same movements and physical locations on the 3x3 grid as the Touch Alone condition and, consequently, provides direct information about the contents of real locations in space. The only difference is whether the identity of the contents at those locations is determined by touch (e.g., feeling a raised X) or by voice (e.g., hearing a spoken X).

The third condition (referred to as Abstract Coordinate) in this experiment requires participants to construct and maintain two-dimensional spatial representations of the tic-tac-toe board without input from any physical objects or locations. In that sense it is more abstract and more amodal than the first two conditions. The board and pieces are referenced by letter (Columns A, B, and C) and number (Rows 1 2 3) coordinates. Thus, the participant could designate the placement of one of his X tokens in the upper-left location by saying "A 1". Likewise, the opponent responds with the coordinate of an open location (e.g., "A 3"). As usual, on his or her turn the participant can review the contents at any coordinate: "What's in B 2?"

In order to control for individual differences a repeated measures design will be used with 30 subjects participating in each of the three conditions: Touch Alone, Touch+Voice, and Abstract Coordinate. The participants will be volunteers drawn from the SFSU Psychology Department participant pool. Each of the three sessions will last about an hour and take place on a different day to minimize carryover effects. The order of the three conditions will be counterbalanced across participants.

A repeated-measures ANOVA with appropriate a priori comparisons will be used to test the following hypotheses. We expect the touch component of the Touch Alone and Touch+Voice conditions to be highly useful in constructing and maintaining the two-dimensional representation of the workspace. If these conditions are not substantially superior to the Abstract Coordinate condition then our underlying conceptual model for the design of our prototypes is undermined. Another important comparison is between the Touch Alone and Touch+Voice conditions as the Touch Alone condition is analogous to our original Braille-based AutOMathic Blocks prototype whereas the Touch+Voice condition is analogous to a second prototype described later that substitutes voice identification for Braille. If the Touch+Voice condition is comparable to the Touch Alone condition, there would be basic research support for the viability of a touch plus voice prototype that can be used by the many visually impaired individuals who are not fluent Braille readers. The literature reviewed above clearly supports our design hypothesis regarding the utility of manipulatives and haptic sensing of spatial location, but as described at the very onset of this section there is a gap between the many decades of basic research on the various sensory modalities and their application to the design of a system such as AutOMathic Blocks. The experiment described in this section should fill that gap.

4 Summary

Through the use of simple games, we expect to get a clearer understanding of the mapping from tactile input to a special representation of mathematics in the visual cortex. As noted, earlier experiments in this domain indicated a strong correlation between tactile input and visual representation (see Muller-Lyer Illusion). In this experiment, a tactile copy of the visual illusion was presented to blindfolded subjects by pressing the solid object on their skin. The results clearly showed that the tactile representation elicited the same responses as the visual representation.

In the current case, we are interested to determine if the same result can be demonstrated in the tactile presentation of math as achieved in the visual presentation. This finding could lead to a whole new set of teaching aids for the visually impaired.

References

1. Karshmer, A.I.: AutOMathic Blocks – An Automated Systems to Teach Math to K-12 Children with Severe Visual Impairment. In: 1st Annual Learning with Disabilities Conference, Dayton, OH (July 2007)
2. Baddeley, A.D., Hitch, G.J.: Working memory: Past, present...and future? In: Osaka, N., Logie, R., D'Esposito, M. (eds.) Working Memory - Behavioural & Neural Correlates. Oxford University Press, Oxford (2007)
3. Berkeley, G.: Philosophical Works, Including the Works on Vision. In: Ayers, M.R. (ed.) Everyman edition. J. M. Dent, London (1709/1975)
4. Dewey, J.: Psychology. Harper & Brothers (1899)
5. Rock, I.: An Introduction to Perception. Macmillan Publishing Co., New York (1975)
6. Casla, M., Blanco, F., Travieso, D.: Haptic perception of geometric illusions by persons who are totally congenitally blind. *Journal of Visual Impairment and Blindness* 93, 583–588 (1999)
7. Millar, S., Al-Attar, Z.: The Muller-Lyer illusion in touch and vision: Implications for multi-sensory processes. *Perception & Psychophysics* 64, 353–365 (2002)
8. Wong, T.S.: The respective role of limb and eye movements in the haptic and visual Muller-Lyr illusion. *Quarterly Journal of Experimental Psychology* 27, 659–666 (1975)
9. Over, R.: A comparison of haptic and visual judgments of some illusions. *American Journal of Psychology* 79, 590–595 (1966)
10. Bean, C.H.: The blind have “optical illusions”. *Journal of Experimental Psychology* 22, 283–289 (1938)
11. Hatwell, Y.: Etude de quelques illusions geometrique tactiles chez les aveugles. *L'Annee Psychologique* 60, 11–27 (1960)
12. Kosslyn, S.M.: Image and brain: The resolution of the imagery debate. MIT Press, Cambridge (1994)
13. Marmor, G.S., Zaback, L.A.: Mental Rotation by the Blind: Does Mental Rotation Depend on Visual Imagery? *Journal of Experimental Psychology: Human Perception and Performance* (2), 515–521 (1976)
14. Carpenter, P.A., Eisenberg, P.: Mental Rotation and the Frame of Reference in Blind and Sighted Individuals. *Perception and Psychophysics* (23), 117–124 (1978)
15. Kerr, N.H.: The Role of Vision in “Visual Imagery” Experiments: Evidence from the Congenitally Blind. *Journal of Experimental Psychology: General* (112), 265–277 (1983)

16. Hampson, P.J., Duffy, C.: Verbal and Spatial Interference Effects in Congenitally Blind and Sighted Subjects. *Canadian Journal of Psychology* (38), 411–420 (1984)
17. Vecchi, T.: Visuo-Spatial Imagery in Congenitally Totally Blind People. *Memory* (6), 91–102 (1998)
18. Barolo, E., Masini, R., Antonietti, A.: Mental Rotation of Solid Objects and Problem-Solving in Sighted and Blind Subjects. *Journal of Mental Imagery* (14), 65–74 (1990)
19. Aleman, A., Van Lee, L., Mantione, M., Verkoijen, I., De Haan, E.H.D.: Visual Imagery Without Visual Experience: Evidence from Congenitally Totally Blind People. *NeuroReport* (12), 2601–2604 (2001)
20. Klatzky, R.L., Lederman, S.J.: Representing spatial location and layout from sparse kinaesthetic contacts. *Journal of Experimental Psychology: Human Perception & Performance* 29, 310–325 (2003)
21. Gaunet, F., Martinez, J.-L., Thinus-Blanc, C.: Early-blind subjects' spatial representation of manipulatory space: exploratory strategies and reaction to change. *Perception* 26(3), 345–366 (1997)
22. Thinus-Blanc, C., Gaunet, F.: Representation of space in blind persons: vision as a spatial sense? *Psychological Bulletin* 121(1), 20–42 (1997)

Audio-Haptic Browser for a Geographical Information System

Limin Zeng and Gerhard Weber

Human Computer Interaction Research Group,
Technical University of Dresden,
Nöthnitzer Straße 46, 01187 Dresden, Germany
limin.zeng@mailbox.tu-dresden.de,
gerhard.weber@tu-dresden.de

Abstract. People who are blind or low vision currently hasn't obtained an effectual solution to access map applications. Although there are existing several paper-based tactile map projects, most of them need additional processing when product of new area of a map. Besides because of the size limitation of paper materials, these kinds of map fail to provide detailed information. In order to improvement accessibility of geographic data, we develop an audio-haptic map browser to access geo-data from an off-the-shelf GIS through a large-scale Braille display. The browser enables to not only maintain lively haptic sensation via raised pins, but also speech out detailed information of each map element stored in the GIS database. Furthermore, in principle it is possible to carry out worldwide map without any additional processing, if the GIS database supports. We employ a novel method, blinking pins, aimed at locating map elements quickly when implementing map search operations. Excepting introduction of our methodologies, we evaluate the system in 2 phases by participation of 4 blind persons. The results of evaluations have been issued in the end.

Keywords: accessible geographic data, audio-haptic interaction, GIS, the visually impaired.

1 Motivation

Geographic Information System (GIS) are applied worldwide due to the location independence of networks and ever growing performance. People enable to look for a street, a hotel, etc., and find the required spot on a vision-based map conveniently. However, for the visually impaired and blind users, it seems to be impossible to benefit from current GIS technology widely, although there are few alternative approaches, like Google Map Textual Interface¹. Even if users with blind and severe low vision utilize screen reader software to access cyberspace, it is impossible to establish some spatial cognition of maps only with speech, independently from visual abilities. A major obstacle is the lack of explicit semantic knowledge linked to spatial geo-information and therefore the inability to represent maps by audio or at least verbal

¹ Google Map Textual Interface (<http://maps.google.com/?output=html>)

explanations. On the other hand, maps made on paper-based materials like swell-paper provide a possibility of building up spatial cognition for the blind, but fail to represent much more detailed and dynamic information, as a result of their limitation of size and characteristic.

In this paper, how to improve accessibility of digital maps and related services from an off-the-shelf GIS are discussed, as well as a haptic map system and its implementation is presented. Based on a novel large-scale refreshable Braille display, the haptic map has the capability of rendering up-to-date geographic data from a Web-based GIS server, which supports Web Feature Service (WFS) specification of OGC². We illustrate the design of multimodal interaction techniques including speech output and gestural input by fingers when exploring maps. With the help of refreshable pins offering haptic perception and audio information of graphical data, the interactive haptic map tries to establish users' spatial cognition comprehensively. Furthermore, an innovative approach for tactile animations through blinking pins has been developed to guide the user while searching.

2 Background

In recent years more and more researchers and projects focus on improving accessibility of maps. Paper-based maps are one of the most common methods to access geospatial data, like via swell paper or Braille embosser. In an international survey of tactile maps [1], microcapsule and thermoform are the two most popular production methods, respectively 68% and 55% of respondents. In addition to their low-cost, both of them could represent tactile discrimination of map elements clearly. Smith-kettlewell's Tactile Map Automated Production (TMAP) Project [2] supports rapid production of tactile street maps of any location in the USA through embossing technologies. It employs a free GIS to render roads and other features in Braille or raised print.

Given the size limitation of only tactile perception map, auditory interaction is a better supplement to render much more information. T3 (Tactile Talking Technology) system [3] is a improvement version of previous TMAP system, enhanced by audio output through the Talking Tactile Tablet (TTT), a portable and inexpensive computer peripheral device as a 'viewer' for tactile graphic materials. When the user touches the surface of the display, which is covered with the embossed map, to explore the area of map, audio feedback is synchronized with the haptic scene. The Blind Audio Tactile Mapping System (BATS) project [4] focuses on the area of spatial awareness provided by visual maps, whose combination of spatial auditory icons and simple tactile vibrations, offers a chance to access additional map information.

With capability of real haptic perception of maps, electronic tactile displays have been employed in recent projects, specifically based on refreshable Braille display. Intelligent Glasses (IG) [5] is aiming at outdoor guidance for the persons with visual impairment, and its tactile map is a simple edge-like representation of the obstacle's location in the scene through an 8 x 8 array of raised actuators. NIST's prototype tactile visual display uses 3600 pins [6] that are raised and lowered to create haptic sensation while rendering graphic documents and maps.

²OGC's WFS standard, <http://www.opengeospatial.org/standards/wfs>

In the meantime, a great deal of new potential materials has been researched to manufacture paperless tactile displays, specifically in micro/nanotechnology. In [7], the Korean researchers evaluate a Braille dot display module equipped thermopneumatic actuators to drive poly-dimethylsiloxane (PDMS) membranes via adjusting the input power of a micro-heater. In [8], a so-called “artificial skin” is demonstrated, which integrates large-scale micro electro mechanical systems (MEMS) based on a stimuli-responsive polymer. The temperature-sensitive polymer can reversibly change its volume by up to 90% through swelling or shrinking in nearly real-time. Swell height is about 0.5mm. Despite this limitation of swollen height, its spatial actuator density of 297 elements per cm², enable to represent objects in high-resolution. It is promising that in the future most users with visually impairments could afford a large-scale tactile display based on novel materials, in order to adopt new applications, specifically geospatial services.

In brief, off-the-shelf GIS embodied mass storage of geographical data, offer a chance to let blind persons understand their living environments via accessible technologies. Among those technologies, a paperless electronic tactile map considered as one of the best potential approach, not only represents dynamic and detailed information, but also enable to establish spatial cognition via combination of real touch sensation and auditory explanation.

3 Methodology

3.1 Infrastructure and Architecture

The BrailleDis 9000 tablet Braille display³ [9] (see Figure 1) can represent tactile graphics on a matrix of 60x120 refreshable pins, in which each module covers 10 (2x5) pins. The 7200 piezo-electric actuators control each pin separately. Touch-sensitivity, low energy consumption and high speed refreshing allow novel interaction techniques. Multitouch sensitivity is based on capacitive sensors, mounted on top of each (2x5) module. Due to its touch sensitive surface, users’ fingers not only explore map directly and conveniently, but the user may also create gestures interaction.

Lines and graphics are rendered similarly on black and white pixel-based monitor on a refreshable Braille display by raised or lowered pins. In our approach, representation of map data from mass-market WebGIS requires geographic data in the format of Geography Markup Language (GML)⁴. Compared to the approach of map image processing, our method is not necessary to require an amount of computer resources, because GML XML-encoding language represents geographic information in the form of text. According to the specification of GML, an item of feature indicates a physical entity, e.g. a road, a building, which describes a list of locations, geometries, and properties. As below, Fig. 2 illustrates the flow-processing of our haptic map system, in which there are 7 main steps.

³ The product is funded by HyperBraille Project, <http://www.hyperbraille.de/>

⁴ GML standards: <http://www.opengeospatial.org/standards/gml>

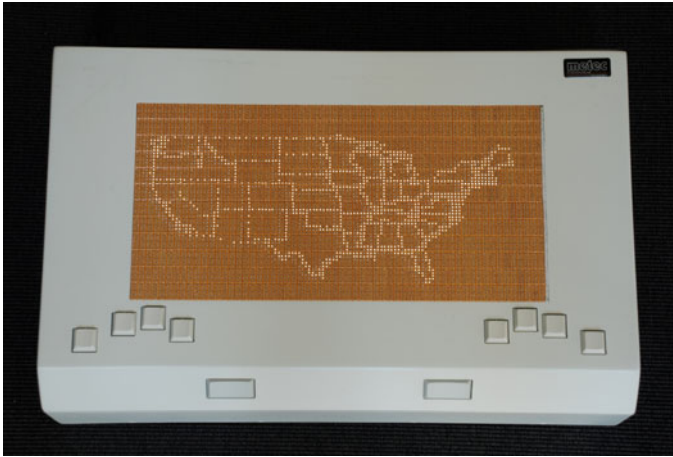


Fig. 1. A US state map (without Alaska) on BrailleDis 9000 (white dots are raised pins)

3.2 Haptic Map Rendering

The task of rendering a haptic map on a large-scale refreshable Braille display is demanding, as the resolution is low compared to a normal screen and only two levels of pins may be utilized to represent lines, textures, text or symbols. Initially the rendering pipeline maps the geographic coordinates into device coordinates and ensures each geo-element is proportionally aligned and corresponds to the correct location. Due to huge amount of information available in a normal WebGIS system, which is suitable for vision representation, but not for tactile output, it is necessary to simplify and filter parts of geographic elements as part of the rendering pipeline. Finally, non-proportional information such as tactile symbols will be inserted where possible.

As shown in Fig. 2, at step 5 all features of geographic elements are determined and map symbols selected according to the various features. For example, when the map element is a bus station, the system looks up the tactile symbol in its map symbol library, and renders corresponding bitmaps.

3.3 Multimodal Interaction

In addition to rendering tactile maps, the haptic map systems supports interactive multimodal interaction. The maps become more readable and accessible through combining audio and haptics feedback. Haptic interaction lets users not only find out “where are you”, but also explore the surroundings via different shaped tactile symbols. Synchronisation of modalities is achieved through sensing contact by one or more fingers or even the palm [10]. By surfing on its surface with one or two fingers the user operates the haptic map for a couple of functions, such as panning, zoom in, zoom out, and verbal explanations.

Blinking pins of an object might be a better method to help users locate points of interest (POI) in a fast manner. The approach of blinking pins is a recurrence of raised or lowered involved pins against the fingertips with tactile sensation. Thus, it is possible to find a blinking element without a slowly scanning the whole of display surface.

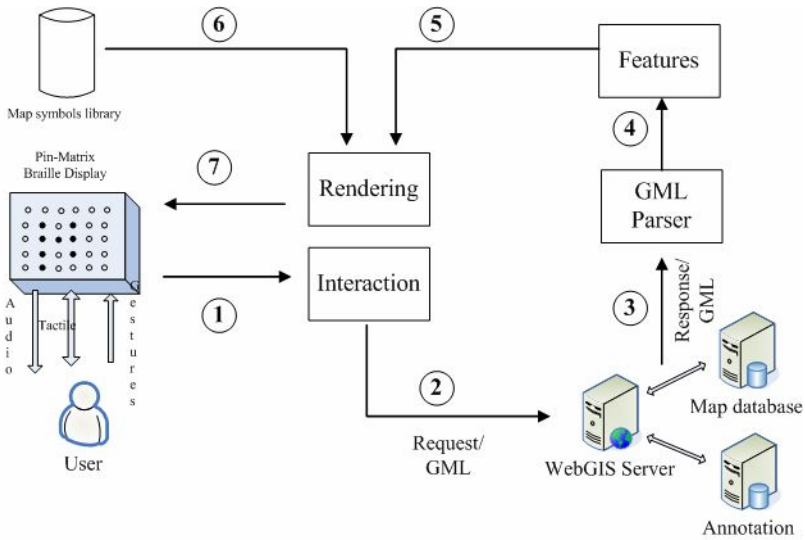


Fig. 2. Data flow by the haptic map system

4 Evaluation

In order to investigate the efficiency of our audio-haptic map browser for the visually impaired to access geographic data, from a GIS, we evaluate the area of our university campus with 4 blind users, among who two are female and two are male. Most of them are familiar with our campus, because one is a worker in university, and two are students in university. The system employs an open source software server, namely GeoServer⁵ allowed users to view and edit geospatial data, as its WebGIS server. The software PostGIS⁶ as a backend spatial database, supports storage of geographic data, as well as implementation of geo-instruction services. Furthermore, we download our university campus map from OpenStreetMap⁷, which is a free collaborative worldwide map platform.

In the evaluation, our tasks focus on estimating accessibility and readability of map elements from a GIS, e.g. roads, buildings, POI, when the subjects explore the tactile campus map. There are 2 phases of evaluation due to improvement of representation technology. Both of them have the same takes.

In the first phase, we received some significant feedback after one blind subject finished the test, that lead us to launch on improving, instead of further evaluations. Among feedback from the subject, the most severe problem was rendering too much geo-information with raised pins at the same time, e.g. roads, buildings, symbols of POI, to understand the map elements where the user was touching. Specifically,

⁵ Geoserver website: www.geoserver.org

⁶ PostGIS website: www.postgis.org

⁷ OpenStreetMap website: www.openstreetmap.org

because of representing the shape of buildings in raised-pin line, it is difficult that the users distinguishes between a building and a road, even if the shape of a building is a closed polygon. Raw geographic data from normal GIS are represented visually and there readability benefits from color, width of lines, icons, etc. but this same effect is not suitable for pin-matrix devices due to its low resolution.

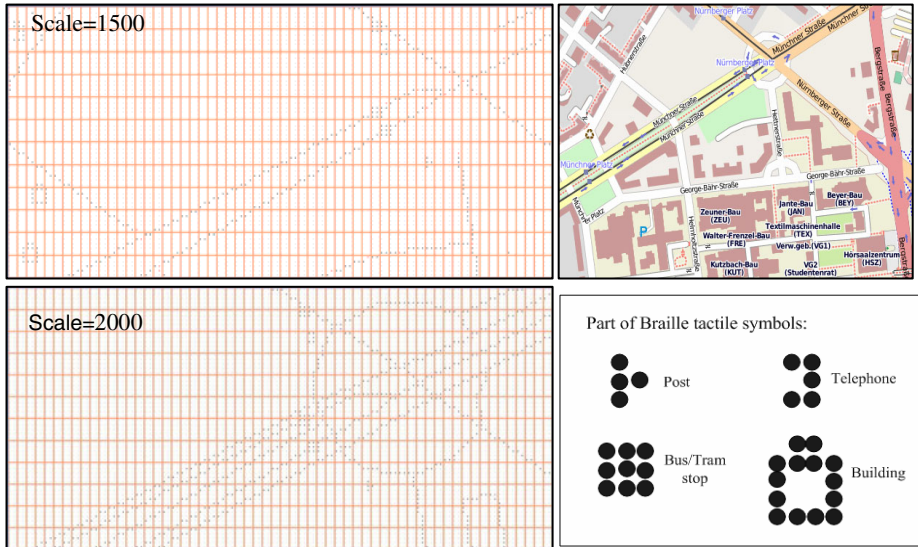


Fig. 3. Different scale of haptic map. Overall-view:Top-left, scale=1500; Detailed-view: Bottom-left,scale=2000; It shows a crossing near highways, tramlines and footpaths (black dot denotes raised pins on Pin-matrix display); top-right image shows campus of TU Dresden from OpenStreetMap; bottom-right image indicates part of Braille tactile symbols.

However, the method of Braille tactile symbols of POI has been accepted by blind pilot users. We design several Braille tactile symbols consisting of raised pins in different combinations to render several POIs, e.g. a bus stop, a traffic light. After a few minutes training of these symbols, the subject not only find out how many symbols on the surface, but also figure out which are of the same kind. Compared to visual map symbols for the sighted and tactile symbols on paper-based map for the visual impaired, Braille tactile symbols help the users know the meaning of a place only with touch.

In the second phase, we invited 4 other blind persons to take part in our evaluation. The system utilizes raised-pins lines only to render roads, while buildings and other map elements are rendered as Braille tactile symbols. Secondly, the system enables to provide 2 kinds of views of the haptic map depended on its scale. One is an overall-view rendering of main roads and POI, the other is a detailed-view with increased number of roads and but including the shape of buildings. Thirdly, ensuing space between elements, which is crucial factor of readability of tactile map, has been

implemented though special algorithms. At last, we redesign a set of Braille tactile symbols systematically, including buildings, bus stop, and other POI.

The study shows that all of them could easily find where a street was, the direction of the street and a crossing consisted of 2 or 3 streets in the overall-view. Their fingers were able to go along the line of a road until its end. These benefited from continuous raised pins and enough space between streets. Additionally, all of them no longer mixed up streets and buildings. Subject confirmed, it is important, the name of the street, the building where her/his finger was pointing is spoken out through pressing one function key on the display simultaneously. 3 of the subjects described this interaction as very easily, and one said “it is indeed easy, but I prefer this information in Braille label”. When we tested zoom in/out with a crossing (see Fig.3) transforming into detailed-view from overall-view, all of them could feel changes and recognize a circle as a sign of crossing, as well as more roads in the area.

After a short time learning the graphical Braille tactile symbols for POIs, maximum lasting 3 minutes, all of them could find some symbols, and know their meanings like a bus station or a telephone booth. However, sometimes participants failed to figure out meaning of a symbol, as they said some of those symbols were too similar to distinguish. Perhaps, after a longer-term of use they can remember all of the symbols. We evaluated also the approach of blinking pins to represent the result of a map search after typing the name of a POI. Blinking pins were highly recommended to locate the result very fast. An approach without blinking required the user to scan the whole map very slowly. Sometimes, they even missed the result.

5 Conclusion and Further Research

For the people with low vision and blind it is difficult to access graphics and in particular maps, thus, we develop an interactive audio-haptic map browser based on a planar tactile display to read a real map from GIS. By implementing a multimodal user interface, users can directly touch the map to enable their spatial cognition, and listen to audio information for requesting more detailed information keeping in the GIS. Blinking pins is a novel method for locating a point of interest or an area fast. Moreover, gesture recognition of fingers is supported to establish a natural interaction for navigating tactile maps. Thereby, these features provide a novel experience and opportunity to access spatial geographic data.

Benefitting from the large-scale Braille display, we can represent a wider area map on it. On the other side, we are developing a smaller and lighter model for portable applications in near future. Future work will take into account feedback during the second evaluation. For instance, when the user just needs to know the shape of a building in the overall-view, there should be a way to show only the shape of the building, without changing into a detailed-view which renders everything in detail. Because there is a huge amount of geo-data in a GIS, a better approach of filtering and simplifying of geo-elements should be evolved in future. Representing a pre-journey route is ongoing via source place and destination, which has been considered helpful by subjects. Moreover, more subjects will be invited to take part in our evaluation to collect feedback and validate effectiveness of our map system. Most of all, we are planning to implement a collaborative platform with the Braille display, in the future, the

visually impaired not only can view geo-data from GIS, but also contribute annotated data to others by themselves.

Acknowledgments. The work is partially based on HyperBraille Project funded by the German Ministry of Economy and Technology (BMW grant, 01MT07004). Thanks to Ursula Weber and other participants. We highly appreciate China Scholarship Council (CSC) for supporting this work.

References

1. Rowell, J., Ungar, S.: The world of touch: an international survey of tactile maps. Part 1: production. *The British Journal of Visual Impairment* 21(3), 98–104 (2003)
2. Miele, J.: Tactile Map Automated Production (TMAP): using GIS data to generate Braille maps. In: CSUN International Conference on Technology and Persons with Disabilities, Los Angeles, CA (March 2004)
3. Miele, J.: Talking TMAP: Automated generation of audio-tactile maps using Smith-Kettlewell's TMAP software. *British Journal of Visual Impairment* 24(2), 93–110 (2006)
4. Parente, P., Bishop, G.: BATS: The Blind Audio Tactile Mapping System. In: Proceedings of ACM South Eastern Conference, GA, <http://www.cs.unc.edu/Research/assist/bats/history.shtml>
5. Velazquez, R., Fontaine, E., Pissaloux, E.: Coding the Environment in Tactile Maps for Real-Time Guidance of the Visually Impaired. In: Proceeding of International Symposium on Micro-NanoMechatronics and Human Science, pp. 1–6 (2006)
6. NIST visual display, http://www.nist.gov/public_affairs/factsheet/visualdisplay.htm
7. Kwon, H.J., et al.: Braille dot display module with a PDMS membrane driven by a thermopneumatic actuator. *Sens. Actuators A: Phys.* 154, 238–246 (2009)
8. Richter, A., Paschew, G.: Optoelectrothermic Control of Highly Integrated Polymer-Based MEMS Applied in an Artificial Skin. *Advanced Materials* 21(9), 979–983 (2009)
9. Völkel, T., Weber, G., Baumann, U., Lüthi, F.: Tactile Graphics Revised: The Novel BrailleDis 9000 Pin-Matrix Device with Multitouch Input. In: Miesenberger, K., Klaus, J., Zagler, W.L., Karshmer, A.I. (eds.) ICCHP 2008. LNCS, vol. 5105, pp. 865–872. Springer, Heidelberg (2008)
10. Schmidt, M., Weber, G.: Multitouch Haptic Interaction. In: Processings of HCI International 2009, San Diego, CA, USA, pp. 574–582 (2009)

A Graphical Tactile Screen-Explorer

Martin Spindler, Michael Kraus, and Gerhard Weber

Human Computer Interaction Research Group
Technical University of Dresden
Nöthnitzer Straße 46, 01187 Dresden, Germany
{martin.spindler,michael.kraus,gerhard.weber}@tu-dresden.de

Abstract. A graphical screen explorer, as it is developed in the HyperBraille project, has different demands on representing information than conventional screen readers and thereby new concepts of interaction, the representation of widgets and the synchronization of multimodal operations become necessary. We describe a concept for complex user interaction with tactile widgets and how to adapt the screen explorer to the requirements of third party applications by the use of Add-Ins.

Keywords: accessibility, blind, tactile graphics display, Braille display, screen explorer, HyperBraille, interaction, extensibility, Add-In.

1 Introduction

The goal of the HyperBraille project is to create a highly modular audio-tactile screen explorer (called HyperReader) supporting a touch-sensitive pin-matrix device with optional speech output allowing blind people better access to and interaction with graphically oriented parts of applications such as business graphics or diagrams [1]. Many toolkits for graphical user interfaces offer software developers sufficient support for drawing complex graphical shapes and flexible control mechanisms, but access to these widgets by blind people requires tremendous efforts – if possible at all. We propose re-designing existing non-visual approaches and the development of appropriate haptic interaction techniques for tactile widgets. In this paper we describe the architecture for a multimodal assistive technology integrating existing screen readers and their enhancement by novel accessibility features related to vector graphics.

Non-graphical screen readers (NGSR) use speech output in combination with a single-row Braille device. Information stems from multiple sources, such as

- Windows-specific accessibility-APIs (UI Automation, MSAA),
- application-dependent Document Object Models (e.g. Word-DOM, PowerPoint-DOM, Browser-DOM),
- direct calls to the Windows-API (e.g. to get the foreground window) and by
- evaluating the data of the video driver.

A screen explorer like HyperReader may use such information but can also be synchronized with any information determined by a NGSr. Current NGSr only allow

disabling them when so-called self-voicing applications take over operating system resources utilizing Braille or speech output.

All available information is stored by the HyperReader in a hierarchical data model (the Off-Screen Model, OSM), that provides further components of the screen explorer, such as the audio- or graphical renderer [2] with additional information resembling the graphical layout of graphical widgets. Additional graphical output on the pin-matrix device as well as the tactile interaction with the application by the user becomes possible by such an approach. Unfortunately, all of these data sources often don't assure sufficient integration and therefore frequently can't provide all necessary information. Hence, there has to be an opportunity to extend the screen explorer depending on the demands of each application.

2 Extensibility Mechanisms in HyperReader

Several available NGSRs, such as JAWS, Window-Eyes or COBRA, are extendable by scripts [3], whereas each of them traces a different approach. Their closer examination reveals three basic types of scripts to distinguish:

- *Default-Scripts* are responsible for the treatment of system events (e.g. focus-, selection-events). Furthermore, they define the basic handling of mouse- and keyboard-interaction and provide information on the management of messages from the usability-APIs and hence form the base for further scripts.
- *Application-oriented Scripts* are based on the Default-Scripts and complement the screen reader with procedures to improve the accessibility of special applications. This may be necessary if the accessibility APIs deliver only insufficient information or an application requires special treatment that exceeds the standard handling of the screen reader.
- *Task-oriented Scripts* can overwrite the default behavior of the screen reader and are used to support application-independent tasks.

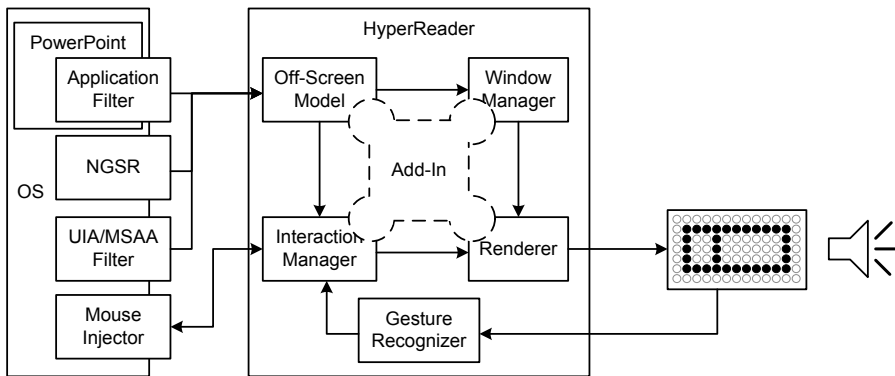


Fig. 1. Extensibility of HyperReader components

HyperReader picks up these concepts and enables software developers as well as sophisticated users to extend the screen explorer to match their needs. The technology applied is based on the Add-In-architecture of Microsoft .NET and uses specified interfaces of the screen explorer to reconfigure and to widen its functionalities [4]. Add-Ins consist of independent DLLs that are loaded at runtime and can be activated or deactivated at any time. In this way, the HyperReader can be extended without modification of the original screen explorer code.

Figure 1 illustrates the Add-In architecture of the HyperReader including its main components. Add-Ins have access to each of them, although only by merely restricted interfaces. This assures that Add-Ins don't have complete access to the screen explorer and only can use a preset of functions. Hence, interfaces for the Add-Ins represent a subset of the internal interfaces of the components. Thereby, an error-free performance of the core functionalities should be guaranteed.

HyperReader supports default tasks, specified by a Default-Add-In and comprises operations such as scrolling, the definition of Braille-window areas or the announcement of the current time. Application-based Add-Ins are capable of overwriting, respectively complementing these operations to adjust the screen explorer to the characteristics of different applications. Finally, rather task-oriented Add-Ins allow the definition of application-independent operation sequences.

3 Interaction with Haptic Application Widgets

With the mechanism described above, HyperReader Add-Ins can define the mapping of visually displayed GUI widgets to their graphic-tactile rendering. Rendering procedures that refer to filtered GUI widgets, stored in the OSM, are defined within Add-Ins and called by the renderer component. These procedures can also include the usage of basic tactile widget shapes, pre-defined by so-called Braille Tactons [5]. Shape and layout of widgets displayed must not only enable the user to easily find and memorize them, distinguish their type and read the meaning of their presented data, but also make him aware of the possible interactions with them. For example, temporarily disabled widgets (often visually grayed out) must be recognizable as such. Widgets might even need to show feedback on actions, like highlighting selected list items.

Besides adapting the graphic-tactile presentational layout of GUI widgets, the tactile widgets also need an adequate concept for user interaction to offer access to their functionality. The spatial nature of tactile HyperReader widgets resembles the two-dimensional character of GUI widgets. In graphical user interfaces, direct manipulation of control elements is state of the art. Breaking the pixel barrier [6] (or at least alleviating its affects) can only be accomplished by also offering the ability of pointing at objects or locations to trigger direct manipulation actions, just like using a mouse or a touch-screen. With NGSR, routing was already possible by pointing at the displayed content, but limited to (mostly single-lined) Braille-text displays [7]. Former research results (with non-blind users) have shown that users tend to use rather unimodal inputs for general action commands whereas interactions on complex spatially displayed content are more likely performed with multimodal input [15]. Apart

from simple keyboard command, HyperReader manages combined keyboard and gesture input operations. With our Add-In approach for offering extensibility to screen reading, software scripts can be written for a certain application that will be executed on specific interactions. These can also be linked to rendered widgets, like mouse event handlers of windows form controls. Pointing operations (e.g. finger tap gestures) can serve different purposes. In HyperReader the diverse action scripts contained in an Add-In can be called by tapping at the very same context while holding a modifier key.

3.1 Exploration and Reading

HyperBraille offers multiple switchable tactile views. These are more abstract than the visual view on screen. For example an outline view is helpful in exploring layout information of GUI and document content, but in this mode no textual information is shown [12]. Figure 2 shows a filtered PowerPoint Slide and its tactile view with HyperBraille in outline view mode. To assist the user while exploring and reading, a pointing gesture at a certain widget can invoke the HyperReader to prompt further properties of the underlying content. The response would be rendered as tactile Braille text in a special status area or read by a synthetic speech renderer, similar to audio-haptic interfaces [8]. For example, the upper big rectangular outlines in Figure 2b and 2c would tell ‘text: Project’, ‘text: Hyper’ or ‘text: Braille’. The application and the HyperReader tactile rendered view are not affected. This exploration technique was tested with 8 blind users who were asked to use this kind of ‘exploration by pointing’ to complete several tasks, e.g. reading names of desktop icons whose tactile forms were just rendered as rectangles. Except for one uncertain user this conception was accepted very well.

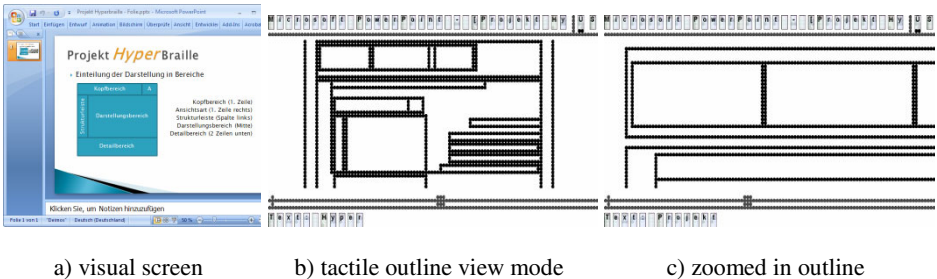


Fig. 2. A filtered PowerPoint Slide (a) and its tactile view (b, c) in outline view mode

3.2 Routing in Tactile Views

Due to the restricted resolution of the HyperReader display, an active control of tactile views is necessary. To browse among displayed content, navigational widgets can trigger actions like page scrolling (tapping on tactile scroll arrow widgets). Other examples are switching between Braille windows by ‘clicking’ on their corresponding list entries and changing between different tactile views by ‘clicking’ in a special

region [12]. Note: the application is not modified. These operations were also tested like described above (see 3.1) and accepted by all 8 users. Using gestures instead of keyboard interaction was generally preferred by 5 users, 3 would prefer it only situational.

3.3 Controlling Applications by Routing

Application-based rendered tactile widgets can be used to control underlying visual application widgets. Therefore, the size and position of the hotspot areas and the widget itself have to be converted from renderer and touch-input position to the real control widget's properties location on screen. To do this, the graphic-tactile renderer maps elements in the off-screen model [9] to rendered tactile widgets (and vice versa). Thereby, simple mouse click interactions can be injected into the system to execute a real mouse click event on the actual control widget after triggering the tap gesture. This could be used for example to enter an edit field by pointing at it in outline view mode, as shown in figure 2.

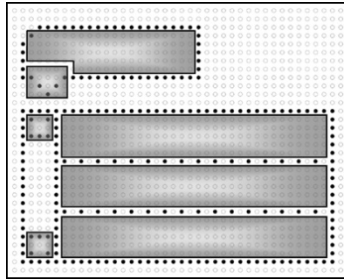


Fig. 3. Multiple routing hotspot areas (gray) of a tactile combo box widget (black dots)

For some widget types, certain areas in the widget can be used as hotspots, since tapping must cause different operations depending on the inner location. For example, Figure 3 shows a possible tactile presentation of a combo box – the user might want to set focus for text input, pop up the drop down list, scroll it up or down or select an item in the drop down list. Even though the conceptual representations of common widget types was well accepted in first user tests with 5 blind experts [10], the interactive functionality is still to be extended.

Former investigations with conventional Braille displays already pointed out how routing mechanisms can be useful for enabling blind people to move the mouse cursor and inject click events [13]. On a touch-sensitive planar tactile display, a fairly high acceptance and usability of routing by pointing at displayed objects can be expected as well.

3.4 Complex Interactions

A popular pointing interaction in GUIs is direct manipulation, namely drag and drop. This compound operation on two objects drags an object from its starting position to a

drop destination while holding a mouse button [11]. But during this procedure not all objects actually affected on the visual screen might fit into the tactile display in HyperBraille at once (due to the lower resolution), so the user needs multiple intermediate browsing operations to find the desired dropping destination. Furthermore, all pins are lowered during gesture recognition for better sensor results due to technical limitations. So the drag and drop operation can't be realized as one gesture without interruption.

As a first approach, a concept was developed, that decomposes the drag and drop procedure into multiple operations. Each of these particular operations injects mouse events ('button down', 'moving' and 'moving and button up') into the operating system:

- To start dragging, a 'grabbing'-operation enters the drag state, similar to copying an object to the windows clipboard. For example, a pointing operation with a timer could be used (a pointing gesture that just takes longer than usual). This would inject a button down to start mouse dragging.
- While in drag-state, the mouse cursor position can be pulled to objects or locations by doing single taps, yet without injecting a click event (avoiding the drop). The user can always query a speech output and status line output to stay aware if he is in drag state and which object or location he is carrying. If the system recognizes an abort event (e.g. hitting the ESC key, but this may differ in some applications), the drag state should be terminated without further mouse event injections.
- When the user has found the desired destination, while in drag-state, the final operation does a drop by injecting a mouse button release event (this could be implemented by a timed tap interaction like for grabbing).

Drag and drop interaction need to be concretely implemented for specific applications in corresponding Add-Ins because its behavior is application-dependent. Along with the definition of a tactile representation of a widget, within the Add-In, it has to be defined:

- whether it is drag-able,
- what drag-able hotspots it has (e.g. adjustment handles or resizable borders),
- for what kind of drag-able objects it could be a valid destination and
- what destination hotspots it has (e.g. for snap-dragging [14]).

If all of this information is provided, the user can be assisted by the screen explorer with audio-tactile hints to successfully complete drag and drop operations similar to a mouse user that recognizes drag-able locations before grabbing and valid destinations while dragging by changing mouse cursor icons. If a user knows where he needs to operate, he can also split the tactile presentation into two HyperBraille-windows to show start and destination side by side (for example to copy files to another folder).

3.5 Technical Implementation of Interactive Behavior via Add-Ins

To distinguish tactile (Braille) reading from gesture input, gesture input is initiated by holding a key while pointing at a location. Another approach is tapping with a second finger while pointing with the forefinger. Since pointing at objects can be used for

different intents, different start keys can be used to address different action script behavior. Recognition and classification of the gesture is triggered by releasing this key. With such a combined multimodal input operation an object is being selected and the reaction is specified by the key used – just like using different mouse keys, clicking on a visual object.

When the pointing gesture is recognized, the interaction manager component in HyperReader queries the tactile graphical renderer to retrieve the object (e.g. an OSM-Element) the user has pointed at. The Add-In responsible for the current application (or at least the global default Add-In) defines the object's tactile rendering and interaction handling. If a specific event handler is defined for this object, it is called. The key combined with the gesture and other session data (e.g. a certain mode like 'dragging') is passed to the handler as parameters to decide how to proceed. Within this handling procedure, synchronization with other running output threads, like continuous speech output, has to be dealt with. Running threads have to be stopped, paused or suspended if they use the same output modalities as the handling procedure does.

If there is no object-specific handler or its execution fails, a default handler is called with the same parameters. It is also possible to call this default script from within an object-specific event handler. For example, the default script could query detail information or inject mouse-clicks on OSM-elements.

4 Conclusion and Outlook

Direct manipulation is a novel concept for blind users. It remains to be evaluated how well complex interactions like the suggested drag and drop concept will be accepted and how effective the given approach can be applied. Yet it appears promising to offer blind people access to functionality in graphic-oriented applications in a similar way to mouse users. Not only widgets, especially graphical shapes like in office or vector graphics applications may become accessible allowing tasks like translating, scaling, rotating and connecting shapes to be solved. Gestures, like striking, pinching, encircling might be useful shortcuts for choosing certain graphical operations. Still, a universal approach for generic applications is impossible because mechanisms like constraints and snap dragging (that are often used in graphical editing environments) need special handling and further multimodal Add-In functionality.

Acknowledgments. The HyperBraille project is sponsored by the German Ministry of Economy and Technology (BMW) under the grant number 01MT07004. Only the authors of this paper are responsible for its content. The Consortium consists of Metec AG, F.H. Papenmeier, Human-Computer Interaction Research Group at the Technical University of Dresden, Institute for Visualization, IZFM institute and IMS-Chips at the University of Stuttgart, the Research Group for Information Processing and Communication at the University of Potsdam and the Blista Marburg as well as T.O.P. GmbH & Co. KG Freiburg.

References

1. Völkel, T., Weber, G., Baumann, U.: Tactile Graphics revised: The novel BrailleDis 9000 Pin-Matrix Device with Multitouch Input. In: Miesenberger, K., Klaus, J., Zagler, W.L., Karshmer, A.I. (eds.) ICCHP 2008. LNCS, vol. 5105, pp. 835–842. Springer, Heidelberg (2008)
2. Kraus, M., Völkel, T., Weber, G.: An Off-Screen Model for Tactile Graphical User Interfaces. In: Miesenberger, K., Klaus, J., Zagler, W.L., Karshmer, A.I. (eds.) ICCHP 2008. LNCS, vol. 5105, pp. 865–872. Springer, Heidelberg (2008)
3. Parente, P., Clippingdale, B.: Linux Screen Reader: Extensible Assistive Technology. In: Assets 2006: Proceedings of the 8th international ACM SIGACCESS conference on Computers and accessibility, pp. 261–262. ACM, New York (2006)
4. Microsoft MSDN Library: Add-ins and Extensibility (April 12, 2010), <http://msdn.microsoft.com/en-us/library/bb384241.aspx>
5. Brewster, S., Brown, L.M.: Tactons: Structured Tactile Messages for Non-Visual Information Display. In: AUIC 2004: Proceedings of the fifth conference on Australasian user interface, pp. 15–23. Australian Computer Society, Inc. (2004)
6. Boyd, L.H., Boyd, W.L., Vanderheiden, G.C.: The Graphical User Interface: Crisis, Danger and Opportunity. *Journal of Visual Impairment and Blindness*, 496–502 (1990)
7. Weber, G.: Reading and pointing - New interaction methods for braille displays. In: Edwards, A.D.N. (ed.) Extra-ordinary human-computer interaction: interfaces for users with disabilities, pp. 183–200. Cambridge Univ. Pr., Cambridge (1995)
8. Gardner, J.A., Bulatov, V.: Directly Accessible Mainstream Graphical Information. In: Miesenberger, K., Klaus, J., Zagler, W.L., Burger, D. (eds.) ICCHP 2004. LNCS, vol. 3118, pp. 739–744. Springer, Heidelberg (2004)
9. Kraus, M., Völkel, T., Weber, G.: An Off-Screen Model for Tactile Graphical User Interfaces. In: Miesenberger, K., Klaus, J., Zagler, W.L., Karshmer, A.I. (eds.) ICCHP 2008. LNCS, vol. 5105, pp. 865–872. Springer, Heidelberg (2008)
10. Albert, K.: Design and Development of a Tactile User Interface (Konzeption und Entwicklung eines taktilen User Interface). Study's Thesis. Human Computer Interaction Research Group, Technical University of Dresden (2008)
11. Buxton, W.: Chunking and phrasing and the design of human-computer dialogues. In: Proceedings of the IFIP World Computer Congress, Dublin, Ireland, pp. 475–480 (1986)
12. Köhlmann, W., Nadig, O., Schiewe, M., Weber, G.: What You Feel is What You Get: Mapping GUIs on Planar Tactile Displays. In: HCI International (2009)
13. Weber, G.: Adapting direct manipulation for blind users. In: Ashlund, Stacey, et al. (eds.) Adjunct Proceedings of INTERCHI 1993, pp. 21–22. Addison Wesley, Reading (1993)
14. Bier, E.A., Stone, M.C.: Snap-dragging. *ACM SIGGRAPH Computer Graphics* 20(4), 233–240 (1986)
15. Oviatt, S., DeAngeli, A., Kuhn, K.: Integration and synchronization of input modes during multimodal human-computer interaction. In: Proceedings of the SIGCHI conference on Human factors in computing systems, pp. 415–422 (1997)

Reading Braille and Tactile Ink-Print on a Planar Tactile Display

Denise Prescher¹, Oliver Nadig², and Gerhard Weber¹

¹ Technische Universität Dresden, Institut für Angewandte Informatik,
Nöthnitzer Straße 46, 01187 Dresden, Germany
{denise.prescher, gerhard.weber}@tu-dresden.de

² Deutsche Blindenstudienanstalt e.V.,
Am Schlag 8/10, 35037 Marburg, Germany
nadig@blista.de

Abstract. Reading characters by fingers depends on the tactile features of the medium in use. Braille displays were often found to be slower than Braille on paper. We study reading Braille on a novel planar tactile display by conducting three reading tests. A first study compares the reading speed on four different devices, namely paper, 40 cell Braille display and two varied conditions on the planar pin device. In a second study we isolate the factor of ‘equidistance’ which is due to the design of our planar tactile display. Our intention is to find out if equidistant Braille can be as fast as standard Braille. Because of the two-dimensionality, the pin device also can show graphics and tactile ink-print. The latter was evaluated with blind subjects in the third study.

Keywords: equidistant Braille, tactile ink-print, blind users, planar tactile display, reading speed.

1 Introduction

In times of graphical multimedia-based applications it becomes more and more difficult for blind people to get equivalent information as their sighted fellows. In the last years, multimodal solutions providing some access to graphical user interfaces for blind people became available. Existing approaches are based upon auditory output of voice or non-speech sound [10], haptic transformations through force feedback [11], or tactile output by Braille. In this paper the planar tactile display ‘BrailleDis9000’ (hereinafter referred to as ‘pin device’) described in [16] is used as input and output medium (see figure 1 a). This tactile device not only allows displaying Braille but also graphics through tactile pins in a matrix of 120 columns and 60 rows at a distance of 2.5 mm horizontally and vertically.

When developing a tactile approach for access to colored graphics or merging any kind of tactile graphics with Braille (see [15]), it becomes apparent that the following problem can arise if text is integrated into graphics. For a continuous presentation of tactile graphics it is necessary that every pin has the same spacing to its neighbors. Because of these equidistant pins, two Braille letters have to be

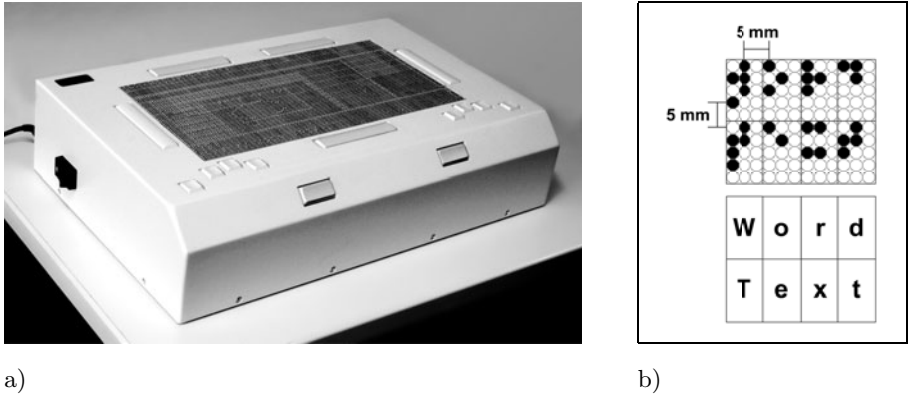


Fig. 1. Planar tactile display (a), 5 x 3 pin-matrix of equidistant Braille letters (b)

separated through a vertical and accordingly a horizontal line of lowered dots on the tactile display in order to be readable. Each letter is then assigned to a 5 x 3 matrix of pins as proposed in [14]. Thus, the spacing among two Braille characters is 5 mm (see figure 1 b), that is 1.5 mm larger than on a standard Braille display. In the following such a Braille font is referred to as ‘*equidistant Braille*’.

Allowing the user an equivalent reading on the pin device, the reading speed should not be adversely affected by equidistance, especially regarding the obvious slower reading speed of Braille readers in comparison to sighted people [3]. Numerous studies can be found addressing several aspects of reading Braille, like hand techniques [6,2], the comparison of reading rates in different tasks [4] or reading Braille in comparison to print readers [17], recognizability of Braille letters [13], and improvement of reading speed [3]. It was also tested how different spacing between Braille dots and between Braille letters influences the readability and therefore the reading speed [7]. But there are no predications for equidistant Braille on a planar pin device yet. For this reason, this paper shall discuss how the reading rate on a two-dimensional equidistant Braille display varies in comparison to a standard Braille device.

Besides displaying Braille, the BrailleDis9000 also allows the output of tactile ink-print fonts. Especially for former print readers and the elderly this could be an alternative reading solution, resembling the barb font (*Stachelschrift*) developed by Wilhelm Klein in 1807. Often these people have difficulties in learning Braille [1]. In a study by Chepaiti [1], seniors could read letters of a raised Roman Alphabet faster and more accurate than standard Braille. This success encouraged us to additionally inspect the use of tactile fonts on a planar tactile display.

In this paper we describe the effects imposed by the tactile medium on Braille reading speed, and discuss initial observations of exploring tactile ink-print.

2 Braille Reading Study

The central purpose of the first study was to compare the reading of equidistant Braille with standard spaced Braille. On the one hand, Braille is read on paper, which is planar, and on the other hand it is read on Braille cells showing only one line at once. In the experiment equidistant Braille should also be presented in a planar way, covering the whole output device and in a single-line modality. A single line standard Braille display may still be read differently if compared to a single line of Braille on a planar tactile display. We call this last mode the ‘Braillex area’. It is the lowermost row of the BrailleDis9000 showing the same information as a standard 40 cell Braille device. Out of this, a total of four different reading conditions arise.

2.1 Method

The study was conducted with 20 blind subjects. After a warm-up phase of 10 minutes on the BrailleDis9000, each subject had to read one of four texts on each of the four reading devices aloud. To avoid learning, order and tiring effects both the texts and the reading media were equally disposed on the four reading tasks. For each device 8-dot Eurobraille was used. At the end of a five minute session the last read phrase was noted. Out of this note, the number of words read was determined and the reading speed in words per minute could be calculated for each device.

2.2 Results

The resulting mean reading speed is shown in table 1. In figure 2 the variability of measured reading speed of all subjects is illustrated for each reading device. The high variation arises from the inhomogeneous group of subjects regarding their reading ability.

The ANOVA revealed a significant effect of the reading device on the reading speed ($F=6.206$, $p = 0.001$). Furthermore, two-sided t-tests show no significant difference on the reading rate between standard Braille display and pin device ($t=2.035$) as well as between pin device and Braillex area ($t=1.932$).

Table 1. Mean reading speed with standard deviation according to the used reading device

Reading device	Mean reading speed in words per minute (wpm)	Standard deviation
Printed Braille (Paper)	58.9	16.3
40 cell Braille device	50.0	12.8
Planar tactile display	45.2	14.0
Braillex area (40 cells)	41.2	10.9

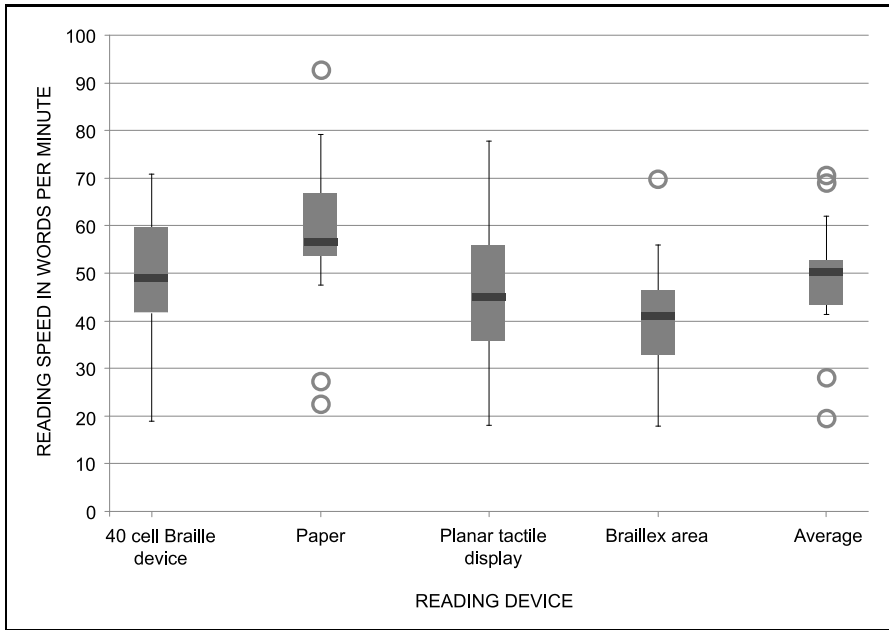


Fig. 2. Comparison of reading speed on different devices

2.3 Discussion

The test has shown that reading on the planar tactile display is not explicitly slower than reading on a standard Braille device. Especially the reading speed of the two slowest readers (identifiable as outliers in figure 2) shows marginal differences on all four reading devices. This suggests that beginners in reading Braille could profit from the major distance between two letters.

However, there were some confounding variables at testing the planar tactile display and the Braillex area, because some pins of the used prototype were damaged. Hence, reading was partially complicated. According to Millar's research, where some subjects spent a lot of time over degraded words, these wrong pins could lead to decreasing factual achievable reading speed by more than 10% [8]. To get more precise results of reading equidistant Braille, a follow-up study was necessary.

3 Follow-Up Study

Isolating the factor of 'equidistance' and, in particular, disabling the considerable confounding variable of incorrect raised pins has been possible by an additional pilot study with four new blind subjects. Hereby not only an initial comparison of different reading rates with standard and equidistant Braille is determined, but also first observations on possible learning effects can be made.

3.1 Method

We choose two subjects with no experience in reading equidistant Braille and two subjects who participated in the project and therefore worked with the planar tactile display for more than half a year. Comparability of the reading material was ensured by printing the text in 8-dot Eurobraille on raised paper in both cases - one in normal space, one as equidistant Braille. After a short warm-up phase, the subject had to read aloud alternating two pages of normal and two pages of equidistant Braille. In all of the four reading phases the needed time was measured and the reading rate was calculated.

3.2 Results

In table 2 the mean reading speed is summarized for every subject. This results in an average reading speed of 84 wpm (normal Braille) and 74.7 wpm (equidistant Braille) for inexperienced subjects, and of 70.1 wpm (normal) and 71.9 wpm (equidistant) for experienced subjects. Thus, it appears that equidistant Braille is not significantly slower than standard Braille.

Table 2. Mean reading speed for standard and equidistant Braille

Subject	Mean reading speed in wpm	
	Standard Braille	Equidistant Braille
1 - not experienced	126.1	109.4
2 - not experienced	41.9	40.0
3 - experienced	68.9	70.6
4 - experienced	71.3	73.3
Average of all subjects	77.0	73.3

3.3 Discussion

It is even seen that the experienced readers are minimal faster in reading equidistant Braille. So we can assume that people can get used to equidistant Braille. Furthermore, this result also confirms our assumption that poorer readers (like subject 2) could benefit from higher letter spacing. In contrast, subject 1 is a very fast reader but requires significantly more time to read the equidistant Braille in relation to standard Braille. Probably this could be explained from the fact that fast readers are very proficient in recognizing Braille letters and therefore need slightly more time because of the longer distance to cover on equidistant Braille.

4 Study with Tactile Ink-Print Fonts

In addition to Braille output, blind users also have the possibility of touching tactile ink-print on the planar tactile display using our screen explorer software.



Fig. 3. Layout view with tactile ink-print

Thus, conventional letters can be displayed on the pin device (see figure 3) and therefore can be perceived tactilely by blind or visually impaired people.

Tactile output on the display can be based on screenshots consisting of pixel data, so the blind user can also make content accessible which is not available using a standard Screenreader yet. For example, text embedded in some graphics can be detected by zooming functions allowing manual resizing of the letters. Furthermore, font changes in standard electronic documents are frequent and are often not detected by blind people [4]. The so called layout view (see [12]) preserves the layout of a document, and users can find, for instance, mistakes in alignment or formatting.

4.1 Method

Identifying the usefulness and the practicality of this tactile view for blind users was the objective of another study with 12 subjects. Half of the subjects had good or medium knowledge of ink-print font in a tactile manner, the other half had minor or no knowledge. The task was to explore a text document, which shows the structure of an invoice. Therefore, the subjects had to work on the planar tactile display showing the content in layout view. The interaction was realized through the buttons of the BrailleDis9000. After navigating and exploring the presented invoice, four of the subjects were presented the same document but with multiple mistakes in formatting and layout. Their additional task was to find these inconsistencies.

4.2 Results

It turned out that the layout view can be a good supplement for blind people, especially for users which cannot read Braille but are able to recognize tactile fonts. Depending on the letter itself and its formatting but also on the tactile ability and the ink-print knowledge of the blind person, a tactile letter can be

identified well from a size of nine pins, which is about 22 mm. Furthermore, the two-dimensional presentation allowed the subjects to perceive spatial relationships within the document. Although 75% of the subjects had some problems in orientation, most of the users rated the layout view positively. Formatting mistakes could be identified easily, but exploring large documents was very time-consuming. In addition, a certain amount of orientation and spatial perception of the user is required to ensure a reasonable usage of the layout view and an efficient way to read tactile fonts.

5 Conclusion

The three conducted studies have shown that users have to get used to equidistant Braille, but they can reach a reading speed comparable to standard Braille devices. Particularly beginners can benefit from faster distinction due to extra spacing between the single letters of the equidistant Braille (see also [9]). Moreover, displaying tactile ink-print can offer completely new possibilities. Primarily persons who became blind at a later age, typically with no Braille skills, can now use tactile information additionally instead of only relying on auditory output.

Based on these findings further research is possible. For example, in the reading tests there was no ascertainment of mistakes while reading on equidistant Braille, although this is an essential criterion for success of a reading device. In detecting layout errors in the layout view a direct comparison of planar tactile displays and standard Braille displays, in respect of how fast and well mistakes can be identified, might be conducted.

Acknowledgements. We thank all blind subjects in Marburg, Dresden and Leipzig for participating in the evaluation and giving us a lot of valuable feedback. In particular, we want to thank Ursula Weber for her support in finding subjects willing to evaluate the display and guiding them through the reading tasks.

The HyperBraille project is sponsored by the German Ministry of Economy and Technology (BMWi) under the grant number 01MT07004. Only the authors of this paper are responsible for its content. The Consortium consists of Metec AG, F.H. Papenmeier, Human-Computer Interaction Research Group at the Technical University of Dresden, Institute for Visualization, IZFM institute and IMS-Chips at the University of Stuttgart, the Research Group for Information Processing and Communication at the University of Potsdam and Blista Marburg as well as T.O.P. GmbH & Co. KG Freiburg.

References

1. Chepaitis, A., Griffiths, A., Wyatt, H., O'Connell, W.: Evaluation of tactile fonts for use by a visually impaired elderly population. *Visual Impairment Research* 6(2&3), 111–134 (2004)
2. Davidson, P., Appelle, S., Haber, R.: Haptic Scanning of Braille Cells by Low- and High-Proficiency Blind Readers. *Research in Developmental Disabilities* 13, 99–111 (1992)

3. Denninghaus, E.: Die Förderung der Lesegeschwindigkeit bei blinden und sehbehinderten Jugendlichen und jungen Erwachsenen. In: *blind-sehbehindert* 116, Zeitschrift für das Sehgeschädigten-Bildungswesen, Verein zur Förderung der Blindenbildung, Hannover pp. 95–100 (1996)
4. Diggle, T., Kurniawan, S., Evans, D., Blenkhorn, P.: An Analysis of Layout Errors in Word Processed Documents Produced by Blind People. In: Miesenberger, K., Klaus, J., Zagler, W.L. (eds.) *ICCHP 2002*. LNCS, vol. 2398, pp. 587–588. Springer, Heidelberg (2002)
5. Knowlton, M., Wetzel, R.: Braille Reading Rates as a Function of Reading Tasks. *Journal of Visual Impairment & Blindness* 90(3), 227–236 (1996)
6. Lorimer, P.: Hand techniques in reading braille; synthesis of spatial and verbal elements of reading. *British Journal of Visual Impairment* 20(2), 76–79 (2002)
7. Meyers, E., Ethington, D., Ashcroft, S.: Readability of Braille as a Function of Three Spacing Variables. *Journal of Applied Psychology* 42(3), 163–165 (1958)
8. Millar, S.: Prose reading by touch: The role of stimulus quality, orthography and context. *British Journal of Psychology* 79, 87–103 (1988)
9. Pester, E., Petrosko, J.: Optimum size and spacing for introducing blind adults to the Braille code. *Re:View* 25(1), 15–22 (1994)
10. Petrie, H., Morley, S.: The Use of Non-Speech Sounds in Non-Visual Interfaces to the MS-Windows GUI for Blind Computer Users. In: *International Conference on Auditory Display*, Glasgow (1998)
11. Ramstein, C.: Combining Haptic and Braille Technologies: Design Issues and Pilot Study. In: *Proceedings of the second annual ACM conference on Assistive technologies*, pp. 37–44 (1996)
12. Schiewe, M., Köhlmann, W., Nadig, O., Weber, G.: What You Feel is What You Get: Mapping GUIs on Planar Tactile Displays. In: *Universal Access in HCI, Part II, HCII 2009*. LNCS, vol. 5615, pp. 564–573 (2009)
13. Schürings, I.: Experimentelle Untersuchung der taktilen Erkennbarkeit und Verwechselbarkeit von Blindenschriftzeichen. *Zeitschrift für experimentelle und angewandte Psychologie* 1987, Band XXXIV, Heft 2, 318–330 (1987)
14. Schweikhardt, W.: Access to the European Videotex-systems by Blind Subscribers. *Communication Systems for the Blind*. In: *Proceedings of a European Workshop in Florence 1986*, Verlag der Deutschen Blindenstudienanstalt e.V. Marburg, pp. 195–202 (1990)
15. Taras, C., Ertl, T.: Interaction with Colored Graphical Representations on Braille Devices. In: Stephanidis, C. (ed.) *Universal Access in HCI, Part I, HCII 2009*. LNCS, vol. 5614, pp. 164–173 (2009)
16. Völkel, T., Weber, G., Baumann, U.: Tactile Graphics Revised: The Novel BrailleDis 9000 Pin-Matrix Device with Multitouch Input. In: Miesenberger, K., Klaus, J., Zagler, W.L., Karshmer, A.I. (eds.) *ICCHP 2008*. LNCS, vol. 5105, pp. 835–842. Springer, Heidelberg (2008)
17. Wetzel, R., Knowlton, M.: A Comparison of Print and Braille Reading Rates on Three Reading Tasks. *Journal of Visual Impairment & Blindness* 94(3), 146–154 (2000)

Enhancing Single Touch Gesture Classifiers to Multitouch Support

Michael Schmidt and Gerhard Weber

Technische Universität Dresden, Institut für Angewandte Informatik
Nöthnitzer Straße 46, 01187 Dresden, Germany
mschmidt@inf.tu-dresden.de, gerhard.weber@tu-dresden.de

Abstract. Recent progress in the area of touch sensitive hardware has accelerated a trend of rethinking classical user interfaces. Applications controlled by gestures have become of special interest. System defined gestures however may prove not to be as intuitive and flexible as preferred by eager users. A same amount of flexibility would be handy for offering special input methods appropriate to users' options and skills. The presented work shows an alternative to system defined gestures even for multitouch input. Extended by a few feature comparisons, existing single touch classifiers are enhanced to support a much wider range of input. The main goal is to provide an easy to use and implement concept of multitouch recognition. Along with the algorithmic design testing results on the classifiers performance and proposals for applications are given.

Keywords: gesture, recognition, classifier, multitouch, recognition rates, user interfaces.

1 Introduction

Gestures as interaction technique are well known and have been applied for a few decades [1], but experienced great attention and interest during the last few years. The continuously growing demand of appliances with gestural input in a wide range of domains (such as mobile phones) seems not to be satisfied in the near future. On the other hand, prototyping of powerful gesture recognition systems is a time consuming task and not the main interest of application designers. Therefore one has to compromise between time and effort spent on the actual design of gestural applications and on the development of algorithms for their proper recognition. Not only developers benefit from ready to use recognition systems, but also researchers. User centered and ergonomic design in the area of gestural interaction would take advantage of more powerful tools, too. Not to forget the actual user who is interested in self-defining some more complex gestures. Of special interest are users who are not able to execute commonly used types of gestures. Predefined types of gestures may be avoided due to different kinds of strategies or habits that result from the usage of other input devices. For instance, tactile displays used by blind people would influence their

approach by a more common usage of both hands to read information and could benefit from a similar and local input method [2]. Teaching gestures to blind people by descriptive methods could be insufficient as well, if no other strategies as demonstrated in [3] are supported. A need for personalized and better memorable gestures could be another reason for defining them oneself.

There are, however, developments that support such needs as described in [4], [5] or [6] to reference a few. But while the first two lack the native support of multitouch gestures, the last one comes at the price of a not easy to handle hidden Markov model approach. A method easier to grasp and implement would be a time-saver in developing prototypes of gestural interfaces. In this work, the authors offer a concept to enhance single touch¹ classifiers² based on templates to support proper multitouch within adequate effort.

2 Improving Gesture Recognition

Preliminary Consideration: If a single touch/stroke³ classifier is capable of supplying a distance measurement in addition to its classification result, multiple interpretations can be balanced. Normally, this classifier will not distinguish temporal or spatial structures of simultaneously or sequentially drawn strokes. To overcome these restrictions an approach similar to Rubine's [5] method is proposed. The original classifier is enhanced to check additional features. In this work, Wobbrock's \$1 classifier [4] was chosen as demonstration object due to the provided and easy to use source code. Alternatively, any other classifier fulfilling our stated prerequisites⁴ would do as well. Another characteristic resulting from this choice is an invariance with respect to rotation, scale and position that should be preserved. Therefore, the chosen methods take such invariances into account without changing the original classifier's characteristics. Modifications that omit invariances are easily achieved and not further discussed at this stage. Note that we will not consider the problem of the Midas touch effect regarding the detection of a gesture's correct duration to avoid triggering wrong gestural commands during haptic contact. We follow the assumption of known starting and end points of each entered gesture, for instance by usage of different modes. Other methods of avoiding this effect such as time slots or special detection of gesture intervals may be applied, but are of no concern here.

Proposed Technique: In a first step the classifier is extended for tap-classification, adding taps in addition to singular touch strokes to a gesture's possible segments called tokens here. Meant to allow for more realistic gestures there came along a

¹ To be more precise, single touch and single stroke would be the appropriate term we substitute for means of convenience.

² With the same convenience we speak of classifiers when referring to gesture classifiers.

³ In this work, input made between the occurrence of a first touch and the first loss of all contacts is defined as stroke.

⁴ I.e. template based single touch classifiers that provide measurements of distances between input and template gestures.

drawback in form of the strong disjunction between tabs and other tokens. Being aware that inconsequently drawn strokes led in small numbers to not correctly classifiable gestures, tabs were included for the sake of practicality. A gesture consisting of a set of tokens, tracked blobs respectively, is now either stored as template or classified against templates already existing. A gesture is classified by comparison with each template and assignment to the one most similar, i.e. with minimal distance. To find the distance to a template, all template tokens have to be matched bijectively to every input gesture's token. Each input token could be classified against every template token by the original classifier. This way a matrix of possible assignments for each input token to a template token plus its distance is retrieved. Assuming distances are normalized to probabilities, getting a global matching probability of template and input is a problem of finding a maximum weighted matching within the bipartite graph this matrix represents. Such matching problems can be solved by the Hungarian method [7] as it is in our implementation though a greedy procedure might be sufficient if token numbers within gestures remain small. See Fig. 1 for an overview of our approach.

So far, gestures are properly classified solely according to their token's shape. But no disjunction between say a pinch gesture and its reverse counterpart may be possible as both types just contain straightened strokes. Neither could a simultaneously drawn approximation sign be distinguished from a sequentially drawn one. More separating features are required for both time and structure

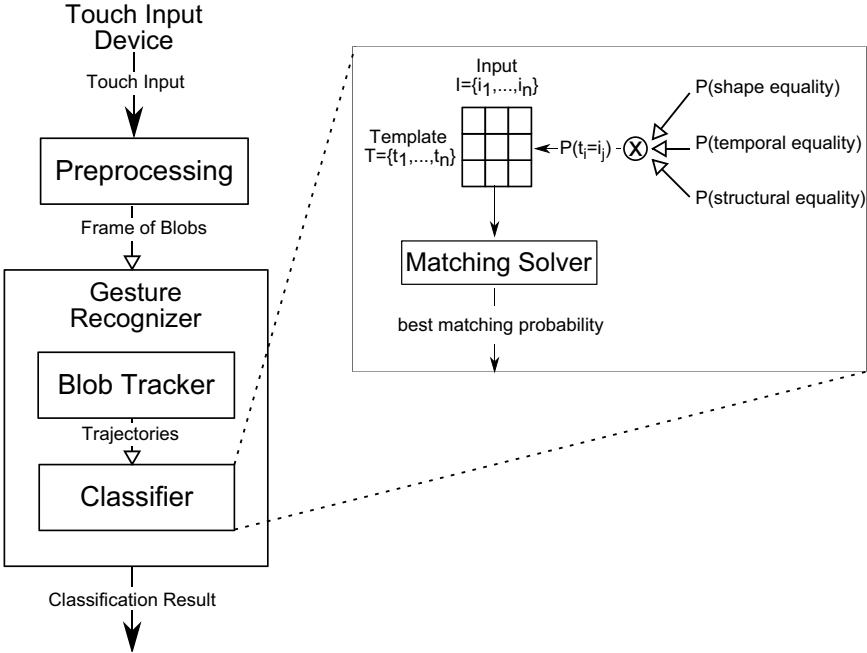


Fig. 1. Gesture recognition architecture

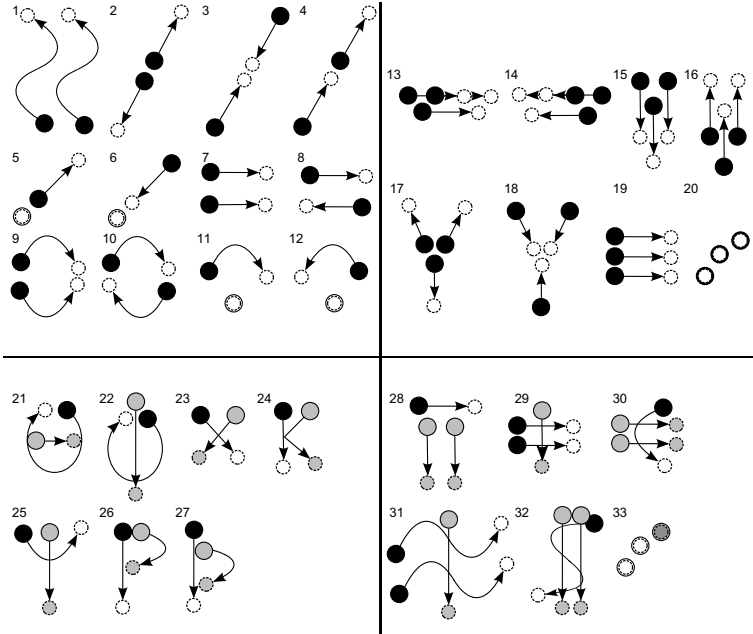
information regarding the inner relations of a gesture's token set. The retrieval of the temporal feature is relatively straightforward and compares a token's starting time to the gesture's starting time relatively to its whole duration. Several more features are needed to take the global structure of a gesture consisting of more than one token into account. These features have to be extracted from a gesture under restrictions the matching method demands. Each feature has to be representative for a single token implying its relation to the gestures remaining part. Structural features should comprise distance and angle information. Five features are chosen to represent a token's posture within a gesture. Three distance criteria regarding positions of a token's starting point, end point and its center of gravity relative to the gesture's center of gravity are computed. For this purpose, the gesture is scaled to fit into a unit square. Furthermore, two criteria for angles are chosen in a similar way. For each token the angles between its starting point or center of gravity and the remaining part's as well as the complete gesture's center of gravity are chosen. The differences of all these characteristics in an input gesture and their counterparts in a template are assumed to follow a normal distribution leading to both temporal correlation and structure matching probabilities as new features. Multiplying the features to the original classifier's token-matching probability leads to the similarity measurement we are interested in. Note that if the order in which gesture strokes are drawn should not influence the classification result, one has to skip computations regarding temporal correlations.

3 Evaluation

Test set-up: To demonstrate the classifier's capabilities, a set of gestures was constructed aiming to contain several types that mostly differ in their inner spatial or temporal features. Therefore, the set should be seen as a tradeoff between potentially usable and challenging gestures. Table 1 shows the testbed. Rows in this table represent the number of strokes or movements without losing touch on the input device⁵, while columns distinguish the number of tokens (i.e. trajectories caused by single contacts) a gesture contains. Filled circles mark starting points of a single finger's touch and dashed ones end points respectively. Each arrow shows the direction to move this finger while missing arrows identify no or very little movement forming already mentioned tabs. Grey symbols depict second strokes, the π gesture (28), for instance, is drawn by making a horizontal one finger left-right movement followed by two fingers dragged downwards. On the other hand, gesture 11 describes a partial circle movement of the index finger around a steady thumb. Note that, as our classifier works with comparisons of token features, any input gesture will only be compared to gesture templates comprising the same number of tokens. The gestures itself consist of few types of tokens keeping them similar enough and avoiding classifications by token numbers and their shapes alone.

⁵ We used a HP TouchSmart tx2 Notebook which is limited to four concurrent recognizable touches.

Table 1. Example set of distinguishable gestures. The set is based on several modifications of swap, pinch and rotation gestures as well as mathematical signs ($=, \neq, \approx, \not\approx$), Greek ($\pi, \theta, \kappa, \Xi, \Phi, \Psi, P, X$) and other (, \$) symbols.



To avoid palliated results due to gestures entered by the author’s hands, four subjects (aged 24-30, two female) were asked to supply five templates for each shown gesture. The gestures were explained with the help of drawings similar to the ones above and hints as to what hand posture is meant to be used were given. Nevertheless, all subjects were allowed to enter their templates as they were comfortable with and would specify them as proper commands. An issue that became apparent at this stage was that some gestures were not comfortable to draw at all. For instance, numbers 13-16 were meant to be entered by forming an isosceles triangle with ones fingertips, the thumb between index and middle finger. Subjects, however, tended towards moving their thumbs outwards resulting in a gesture’s shape similar to number 19. Another interesting feature is the heterogeneity of gestures produced by different subjects. It is clear that there is a greater possible variance not only due to a more difficult explanation of multitouch and multistroke gestures, but to enhanced degrees of freedom in drawing them. A comparative evaluation of classifiers could profit from user centered design of gestures as shown in [8].

The set retrieved was randomly divided into actual templates and test input gestures. Table 2 shows per gesture as well as averaged results of our testing routine.

Table 2. Per gesture and average recognition rates for all and singular subjects. Coloured rows depict overall worst (light) and best (dark) results for specific gestures.

Nr.	All Subjects					Subject 1		Subject 2		Subject 3		Subject 4	
	n	1	2	3	4	5	1	2	1	2	1	2	1
1	0,70	1,00	0,90	0,90	1,00	1,00	1,00	1,00	1,00	1,00	1,00	0,75	1,00
2	0,10	0,80	0,90	1,00	1,00	1,00	1,00	1,00	1,00	0,75	0,67	0,75	0,67
3	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
4	0,80	1,00	0,90	1,00	1,00	1,00	1,00	1,00	1,00	1,00	0,33	1,00	1,00
5	0,90	0,70	1,00	0,90	1,00	1,00	1,00	1,00	1,00	1,00	1,00	0,75	1,00
6	0,90	1,00	1,00	1,00	1,00	1,00	1,00	0,75	1,00	1,00	1,00	1,00	1,00
7	0,50	0,80	0,70	0,80	0,60	0,75	0,67	1,00	1,00	1,00	1,00	0,75	1,00
8	0,30	0,40	0,10	0,50	0,40	0,50	0,67	1,00	1,00	0,75	0,67	0,25	1,00
9	0,80	1,00	1,00	0,90	1,00	1,00	1,00	0,00	0,33	0,00	0,33	1,00	1,00
10	0,80	0,80	1,00	1,00	1,00	1,00	1,00	0,75	0,67	1,00	1,00	1,00	0,67
11	0,20	1,00	0,90	0,90	0,90	1,00	1,00	1,00	1,00	0,00	0,67	1,00	0,67
12	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	0,50	0,67	1,00	1,00
13	0,50	0,90	0,70	0,80	0,90	1,00	1,00	1,00	1,00	0,75	1,00	1,00	0,33
14	0,80	0,80	0,30	0,80	0,70	1,00	1,00	1,00	1,00	0,75	1,00	0,75	0,67
15	0,60	0,30	0,80	0,80	0,80	1,00	1,00	1,00	1,00	0,75	1,00	1,00	1,00
16	0,40	0,50	1,00	0,80	0,70	1,00	1,00	1,00	1,00	0,75	1,00	0,50	1,00
17	0,60	0,80	0,40	0,80	1,00	1,00	1,00	1,00	1,00	0,50	1,00	0,75	0,00
18	0,70	0,80	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	0,75	1,00
19	0,50	0,70	0,70	1,00	0,50	0,75	1,00	1,00	1,00	0,75	0,67	1,00	0,33
20	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
21	0,70	1,00	0,90	0,80	0,90	1,00	1,00	1,00	1,00	1,00	1,00	0,75	1,00
22	0,80	0,90	0,60	0,90	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
23	1,00	0,90	0,90	0,80	0,90	1,00	1,00	1,00	1,00	1,00	1,00	0,75	1,00
24	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	0,75	0,67	1,00	1,00
25	1,00	0,90	1,00	1,00	1,00	1,00	1,00	1,00	1,00	0,50	0,00	1,00	1,00
26	1,00	1,00	1,00	0,90	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
27	1,00	1,00	0,90	0,90	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
28	1,00	1,00	1,00	1,00	0,90	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
29	0,80	0,60	0,80	0,40	0,70	0,50	0,67	1,00	1,00	0,50	0,67	0,75	0,67
30	1,00	0,90	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
31	0,80	1,00	0,80	1,00	0,90	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
32	1,00	1,00	1,00	1,00	1,00	1,00	1,00	0,50	1,00	0,75	1,00	1,00	0,67
33	1,00	0,90	0,90	1,00	0,80	1,00	1,00	0,25	1,00	1,00	1,00	0,75	1,00
∅	0,76	0,86	0,85	0,90	0,90	0,95	0,97	0,92	0,97	0,81	0,86	0,88	0,87
∅	0,66	0,78	0,83	0,82	0,82	0,93	0,98	0,82	0,90	0,77	0,82	0,78	0,87

Results: To give a full picture of the classifiers behavior regarding different influences, Table 2 comprises more detailed results. The first five columns show results gained by ten input gestures per type classified against one to five templates per type out of the entire set. The following eight columns provide information per subject for one and two templates of each gesture out of the given five with four

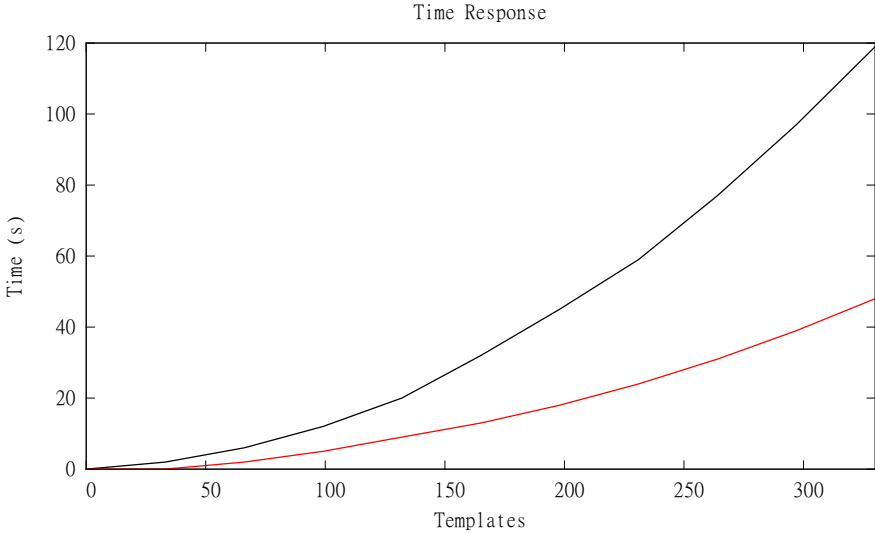


Fig. 2. Classifiers time response with (top) and without (bottom) usage of the single touch classifier for token's shape comparison

and three inputs respectively. The last line shows average results gained by omitting the single stroke classifier and therefore tokens' shape equality computations completely⁶, incorporating no comparison of two tokens' shape equality. These values show the separating performance of solely introduced features. It seems that the small feature set combined with the proposed assignment method is sufficient to handle many multistroke gestures by itself with only a few templates. Some decline in the classifiers performance at growing numbers of templates seem to indicate singular differing gestures chosen as templates by the random selection. This becomes apparent for number 9 as it has mostly good results, but few outlier dropping overall performance in contrary to number 8 that achieved low values more often than not.

Figure 2 shows the time response of our current implementation. The measurement was done once more with and without usage of tokens' shape discrimination for one to ten templates of each gesture presented above. Time was taken for input classification of 33 gestures, again one of each type. As there is not much benefit of increasing template counts per gesture beyond the range of five, the shown behaviour is considered adequate.

4 Conclusion and Future Work

An easy to extend set of recognizable multitouch gestures opens a lot of possibilities. Besides improvement of users' acceptance of an application, it is a time

⁶ In this case it is the \$1 classifier.

saver for designers. Other positive side effects are the options offered in the field of sketching. This way tactile feedback enhanced touch input devices like [9] provide input options for blind or otherwise handicapped people, for instance, of mathematical symbols. The work at hand presents a concept for modifying existing single touch classifiers to support multitouch, but is by no means at an end. Two apparent things come in mind while considering this approach. First, regarding the technical aspect, one may analyze weighting strategies regarding the new temporal and spatial features, more features or speedup possibilities. Second, there is the issue of evaluating multitouch gesture recognizers in general. Leaving subjects to draw a given set of gestures may not be an optimal solution as this approach lacks practical relevance and therefore leads to biased results. While involving users in the design of gesture sets used for evaluation can be appropriate, using gestures specified by them in real world applications can be an even better choice.

References

1. Buxton, W.: Multitouch Systems That I Have Known And Loved (2010), <http://www.billbuxton.com/multitouchOverview.html>
2. Sturm, I., Schiewe, M., Köhlmann, W., Jürgensen, H.: Communicating through gestures without visual feedback. In: PETRA 2009: Proceedings of the 2nd International Conference on PErvasive Technologies Related to Assistive Environments, pp. 1–8. ACM, New York (2009)
3. Schmidt, M., Weber, G.: Multitouch Haptic Interaction. In: UAHCI 2009: Proceedings of the 5th International on Conference Universal Access in Human-Computer Interaction. Part II, pp. 574–582. Springer, Heidelberg (2009)
4. Wobbrock, J.O., Wilson, A.D., Li, Y.: Gestures without libraries, toolkits or training: a \$1 recognizer for user interface prototypes. In: UIST 2007: Proceedings of the 20th annual ACM symposium on User interface software and technology, pp. 159–168. ACM, New York (2007)
5. Rubine, D.: Specifying gestures by example. SIGGRAPH Comput. Graph. 25(4), 329–337 (1991)
6. Damaraju, S., Kerne, A.: A Gesture Learning and Recognition System for Multitouch Interaction Design. Poster session presented at: The Future of Interactive Media Workshop on Media Arts, Science, and Technology (MAST) (January 2009)
7. Kuhn, H.W.: The Hungarian method for the assignment problem. *Naval Research Logistics Quarterly* 2, 83–97 (1955)
8. Wobbrock, J.O., Morris, M.R., Wilson, A.D.: User-defined gestures for surface computing. In: CHI 2009: Proceedings of the 27th international conference on Human factors in computing systems, pp. 1083–1092. ACM, New York (2009)
9. Völkel, T., Weber, G., Baumann, U.: Tactile Graphics Revised: The Novel BrailleDis 9000 Pin-Matrix Device with Multitouch Input. In: Miesenberger, K., Klaus, J., Zagler, W.L., Karshmer, A.I. (eds.) ICCHP 2008. LNCS, vol. 5105, pp. 835–842. Springer, Heidelberg (2008)

User-Interface Filter for Two-Dimensional Haptic Interaction

Wiebke Köhlmann¹, Francis Zinke¹, Maria Schiewe², and Helmut Jürgensen^{1,3}

¹ Institut für Informatik, Universität Potsdam, August-Bebel-Straße 89,
14482 Potsdam, Germany

² F. H. Papenmeier GmbH & Co. KG, Talweg 2, 58239 Schwerte, Germany

³ Department of Computer Science, The University of Western Ontario, London,
Ontario, Canada, N6A 5B7

Abstract. General principles for the design of user interfaces for systems software are discussed, which take into account the constraints imposed by special user groups like visually impaired or blind persons and by special displays like haptic displays or very small screens.

1 Background

A haptic interface, possibly supplemented by sound, for standard computer applications like office tools would make it significantly easier for a blind person to work in an integrated computer-based environment. Screenreaders combined with Braille displays and standard or Braille keyboards provide a modest, but important partial solution: They can provide contents information and allow for (mostly) command-based interaction; they cannot really provide layout information nor mouse-type touch interaction. The newly developed planar, touch-sensitive and refreshable haptic display, the BrailleDis 9000, provides the technical means to build a haptic interface which shows both contents and layout information present on a computer screen and to manipulate this information directly. This is the ideal view; in reality, in building such an interface one encounters many obstacles: these are fundamental, that is, not just results of shortcomings in the technology.

We present recommendations for principles of interface design which simplify using common application programs for blind users and many other types of users as well. These recommendations are based on our experience with building haptic interfaces for WORD and INTERNET EXPLORER.

The BrailleDis 9000 (see [20]) is developed as part of the HyperBraille® project [8]. The display consists of 60 rows of 120 dots each at roughly the usual Braille dot distance horizontally and vertically, which can be raised or lowered; moreover the positions of some or all fingers on the display are available as input through sensors. The display can be used to present both text and graphic as output and to accept gestures as input (e.g. [7,19]). The BrailleDis 9000 supercedes an earlier, less flexible display of this kind, the METEC DMD 120060. To our knowledge there are currently no other such displays. For a comparative

study of various display technologies see [3]. To represent planar information, even graphics, to a blind person, various types of hard copy are normally used. These can be made partially interactive in combination with a pointer-sensitive device like the Talking Tactile Tablet [10]. For a comprehensive survey see [15]. On hard copy a greater resolution, both in the plain and vertically, can be achieved than on the BrailleDis 9000; this limitation of the BrailleDis 9000 may be weakened by future development of technology. The limit for the resolution is, however, essentially not a technical one, but to a large extent a consequence of the limitations of the sense of touch.

2 Screen Interfaces

The typical application programs for which haptic interfaces are developed in the HyperBraille project include, but are not limited to, INTERNET EXPLORER, WORD, EXCEL, POWERPOINT and their counterparts in other operating environments. The BrailleDis 9000 allows for planar rendering of information and interaction by gestures. Mouse actions could be performed by gestures; the screen layout could be rendered as a haptic image. These features can be combined with acoustic communication. Compared to other devices the following is gained in principle:

1. In contrast to speech output and Braille displays, *planar* information is available directly through touch and input through gestures is possible.
2. In contrast to static planar haptic displays, like the Talking Tactile Tablet [10], changes of information can be displayed without delay. Moreover, the display window can move without special preparation of information material.

Many basic issues as to how blind persons comprehend planar information still require detailed investigation (see [2]). However, there seems to be a consensus regarding the following issues: non-textual planar information is understood bottom-up; scrolling, zooming, window displacements can be confusing; hand and finger movement should be small-scale; features must be easy to understand and well separated. This list is certainly not complete, but indicates some of the difficult requirements. A detailed general list of recommendations for haptic interaction is provided in [9]. In this paper we do not discuss specific implementation problems we encountered, but focus on interface design issues arising from the potential and limitations of the new display.

The main goal is as follows: A source provides information to be read, manipulated, acted upon etc. The information should be identical for all users in all relevant aspects; the same actions should be available to all legitimate users with little difference regarding performance or stress etc. For an application program like WORD this implies for instance: not only the contents of the document has to be available, but also its structural information, like layout or print style; manipulating the document like changing text or layout must be possible in essentially the same way for all types of users; thus, in particular, user interaction devices like menu items, buttons etc. ought to behave in the same way. A point expressed often by blind computer users is that they wish to get information

in the same *form* as the sighted users – as a reason it is emphasized that this will facilitate the interaction between blind and sighted persons (see, e.g. [1]). A slightly different view is proposed in [4,14].

3 Presenting a Screen

It is impossible to present the complete content of a screen (or of a window) on the BrailleDis 9000. Even to present just the frames of the areas of a full-screen window of WORD by lines made of Braille dots requires an area of roughly 2 m×1 m: No text would be provided; the menu elements would only be indicated by small rectangles without information; of course there would be no colours. To add information would only increase the size. Using a smaller distance between dots might reduce the size a bit. Even if such a display could be built and were affordable, it would not be usable. It is, therefore, inevitable that only a comparatively small part of the screen view is rendered on the haptic display – the haptic window within a screen window. For a sighted user this problem arises when the screen window is not large enough; scrolling or enlarging the window is a partial remedy. The haptic window cannot be enlarged, nor is scrolling a desirable option. In [18] a structured approach is described which has been adopted in the HyperBraille project. We explain the idea briefly.

The haptic window is subdivided into four regions as shown in Figure 1. The *header region* displays information about the application, like status, title, menus. The *body region* is used for contents of the screen window. The purpose of the *structure region* is to provide information about items in the body region. The *detail region* displays information about items currently selected like status of a menu item or co-ordinates of a cursor or colour or font etc. Regions can be maximized or minimized and can have different views. The layout is designed to support bimanual operation on the BrailleDis 9000 including input by gestures.

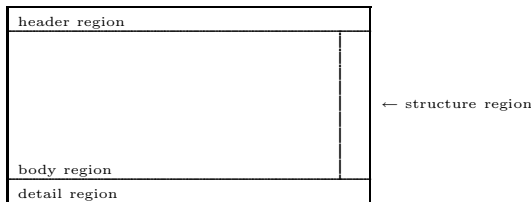


Fig. 1. Regions of the haptic window [18]

4 User Interfaces of Standard Software

Our recommendations concern interaction techniques. These will affect mainly the header and detail regions of the layout described for the BrailleDis 9000 or, more generally, the logical structure of an interface. We explain the recommendations using the example of WORD 2007. We also refer to L^AT_EX as a

radically different paradigm (see [12]). These systems serve to illustrate specific points to save space otherwise needed for lengthy general explanations. Similar observations could be made about the user interfaces of many other software systems. We assume the reader to be familiar with such interfaces and limit our descriptions to the important issues.

With no special adjustments, upon starting WORD the user gets an opening screen as follows:

- In the top left corner there are there are four selections: (1) file options like *new, open, save, save as, print, prepare, send, publish, close*; (2) shortcuts to saving, undoing, re-doing, and printing. The names of some of the file options are not very intuitive; most file options have several sub-items.
- In the bottom right corner there are 5 choices of the document view and a zoom slide. These items are also available through the *view* menu.
- On the left, in the main screen section one can opt to display or hide the clipboard.
- The top row allows for 8 menu choices: (1) *Home*, (2) *Insert*, (3) *Page Layout*, (4) *References*, (5) *Mailings*, (6) *Review*, (7) *View*, and (8) *Acrobat*. Each of these tabs makes several groups of choices available. For example, in the case of *Home* they include font, font size and font style selection, character positioning, ‘special’ characters, paragraph and list shapes, selection of heading and text styles, and a few simple editing tools.

This is confusing on a screen of the usual size, but even more so when there is insufficient space: not only on haptic displays, but also on hand-held computing devices. In a typical document very few of these options are actually used. Usually the appearance of a document is determined by its logical structure together with a few initial design decisions. We examined documents of various types, prepared using WORD and L^AT_EX, for the usage of special features. In general, once the document type and its style parameters have been selected, very little variation and very few special features are actually sought. Because of the differences in the internal representations of documents, obviously it is much harder to find user patterns in documents prepared in WORD than in L^AT_EX.

- In L^AT_EX the main selections are made through the parameters of the commands `\documentclass` and the `\usepackage` and through environment switching either by simple commands with parameters or by explicit environment selections. The complete capabilities of T_EX are available, but not immediately visible, to the author of a document at all times. However, only in extremely rare cases is there a need to resort to them.
- WORD offers an overwhelming and nearly unstructured selection of templates from which to choose. Neither are the templates well documented, even their existence is rarely mentioned, nor do most of the manuals explain them as a key feature or how to create templates. Very little can be found on templates in the official Microsoft document [13]; more can be found in the *Missing Manual* [5], but their use is hardly encouraged. Templates are similar to forms: the author enters the text in the appropriate areas. The final appearance of the text is determined by the settings for that area like in an environment. Unless the author takes explicit steps to remove or hide the

panoply of menu choices, the full set of menu items is available to override the template settings at all times with sometimes unpredictable effects. The usage of templates, like forms, relies heavily on visual features.

The user interface provided by WORD violates well-established rules. Essentially the same can be said about the other OFFICE tools. In this specific context, the following guidelines for improvement are nearly obvious:

- There should be a very small selection of default configurations with few additional parameters (*styles* or *templates*).
- Options need to be structured such that those which have been de-selected by the default configuration or which are used rarely are hidden, but can be found with some effort. Options which are used frequently should be found easily and should be simple to use.
- Abstraction regarding object features is required to focus on the meaning rather than the appearance. The principles underlying the design of L^AT_EX, in particular Version 3, could be used for guidance.
- The interaction system needs to learn user preferences and usage statistics.
- One needs to distinguish between the creation of a new document and the manipulation of a pre-existing document; in the former case, the user can select the style in the beginning; in the latter the system needs to infer the relevant options from what has been used in the document.
- No specific ‘normal’ user model should be imposed. The configuration of the interface would be achieved mostly through definable styles supplemented by dynamical adaptation.

This list is not complete. To show that such an approach is feasible we list a few suggestions. The following items of the WORD user interface can be moved to some remote background, once a document style has been selected:

- most of the rarely used items on the main menu;
- all font selections including font size, bold, slanted, etc.;
- paragraph shapes, spacing and styles to be replaced by standards;
- style changes as the style is determined by the initial style selection for the document;
- ...

We cannot complete the details of the list due to space limitations for this paper. The reader will have understood the point from these examples. More than 95% of the options provided by the WORD interface can be moved without any serious usability problem, as most of the options would either be invoked by the environment anyway or never be used.

For the design of a user interface this implies that a far leaner menu will serve the purpose at a greater convenience for the user. Moreover, a leaner menu fits more easily into limited space like that of the haptic window or a cellular phone. It is far better usable. For many standard systems, versions exist which have been scaled down to be used on devices with very small screens. Unlike these scaled-down systems, our approach preserves all the capabilities of the application.

The new modules will behave as a filter for interaction systems different from the original ones; the methods of dealing with menus in HyperBraille are consistent with this approach as should also be those of the simplified interaction system

of the Dolphin Guide [6]. Beyond easing the constraints imposed by small windows on interactions, by moving much of the unused interaction clutter into the background, the interaction process is made more efficient and less stressful for most types of users. This approach does not distract from the WYSIWYG principle and its haptic counterpart; instead it makes it easier to get what one wants.

5 Interaction Modes

Structuring and shrinking the feature space is a first essential step. With fewer menu items to choose from, the interaction is simplified, the space required is reduced and, nevertheless, the complete functionality is preserved. For small screens with a pen-based interface this may be sufficient. On the haptic display, using the layout shown in Figure 1, there may still be too little space for a well separated placement of the interaction items, even when they are rendered with minimal detail and using specially designed haptic icons. While relying on haptic interaction completely would be feasible, added sound output further simplifies the presentation of the choices. This can be combined with a basic input gesture like tapping a specified spot on the display when the appropriate item is spoken or using voice input. As lists of items will be short for standard selections, this approach, which resembles the choice of menu items on a touch-tone telephone, would be at least as efficient as and hardly any more boring than the pointing selection from a list.

The Dolphin Guide [6] provides a simplified voice-supported interface. It is easy to use, but restricts the access to functionality. For document processing, in particular, it offers the choice of a few templates, the input of which it controls through prompts. In essence, it seems that Dolphin Guide achieves, in a limited way, what our systems architecture is designed to do in full.

It is common knowledge that the combination of touch and sound can improve the interaction compared to when one relies only on one of these modes. Such an integrated approach has been taken successfully in a wide range of scenarios like reading structured documents and tables [17] or exploring diagrams [4,11,14]. It is also supported by the HyperBraille architecture. Whether adding voice output to the interfaces of small pen-based devices would also be beneficial seems not to have been investigated.

The following issue recurs as being contentious in discussions of the way information is presented to blind individuals: To what extent should the rendering of the information for the blind person resemble that for the sighted person. The answer depends on goals, user background and many other parameters. We have seen proposals for a radical deviation of the appearance the haptic representation of information from the visual one; we have also encountered strong opinions advocating extremely close resemblance. The study [1], for instance, reports that a majority of blind computer users would prefer a strict resemblance of the haptic image to the visual image. One of the main problems with this study and similar ones (and opinions expressed by individuals) is that there usually is too little empirical experience of the test subjects to make an informed decision. Admittedly,

an extremely important point is that people with different abilities need to be able to communicate not only via some means (like a computer) but also about those means. For the user interface to an application program this implies, that, at least on the level of metaphors, there should be a common ground for all users.

Our proposal, while motivated by the constraints of haptic interaction, is equally adequate for sighted users of small displays (like those of cellular phones) and for sighted users of ‘normal’ screens. The interaction is streamlined, but no functionality is lost.

Much needs to be done about instruction manuals. We have consulted many different manuals, among these about 25 for OFFICE 2007 and, especially, WORD. Even manuals which claim to address the needs of a special group like [16] do not tell one how to achieve a task, but dwell on superficial details. Online explanations, like those in the WORD interface, are rarely helpful in a complicated situation, thus not a substitute for a manual, but are extremely useless when the display channel does not support the simultaneous display of the problem at hand and the help information. On a visual display window switching is an unsatisfactory solution; on a haptic display, it is utterly confusing. What is called for, is a well-written manual, not just a help system.

6 Conclusions

Our paper addresses general issues of interface design. It is motivated by the constraints and necessities of haptic or multi-modal interfaces for blind users of standard computer applications. The proposed model is applicable to computer usage in general, but particularly important for interfaces on computers with very small displays. We advocate removal of menu cluttering in favour of semantic-based choices. This leads to far leaner interaction systems without restricting the access to system functions; it advocates a logically structured interaction based on structured goals rather than on minor tasks.

Acknowledgements

This research was supported in part by a grant from *Bundesministerium für Wirtschaft und Technologie, Germany*, through the HyperBraille project, grant number 01MT07003, and also by a grant from the Natural Sciences and Engineering Research Council of Canada.

References

1. Report of Tactile Graphics Sub-Committee Part 3. Canadian Braille Authority, English Braille Standards Committee, July 10 (2003)
2. Breidegard, B.: Att göra för att förstå– konstruktion för rehabilitering. Doktorsavhandling, Lunds tekniska högskola, Certec, Avdelningen för rehabiliteringsteknik, LTH Nummer 2 (2006)
3. Brückner, B., Schönecker, A.: Technologiestudie zu Brailledisplays, Report, Fraunhofer Institut für keramische Technologien und Systeme (IKTS), Dresden, Germany (March 2010)

4. Eramian, M., Jürgensen, H., Li, H., Power, C.: Talking tactile diagrams. In: Stephanidis, C. (ed.) Proceedings of HCI International 2003, Universal Access in HCI, Inclusive Design in the Information Society, vol. 4, pp. 1377–1381. Lawrence Erlbaum Associates, Mahwah (2003)
5. Grover, C.: Word 2007: The Missing Manual. O'Reilly Media, Inc., Sebastopol (2007)
6. Guide, easy to use, easy to learn; guide user manual. Software Express Limited, p. 135 Version 4 (2007)
7. Gutschmidt, R., Schiewe, M., Zinke, F., Jürgensen, H.: Haptic emulation of games: Haptic Sudoku for the blind. In: Proceedings of PETRA 2010, Samos, Greece, June 23–25, p. 8. ACM Press, New York (2010)
8. HyperBraille[®], <http://www.hyperbraille.de/>
9. ISO 9241-9xx: Ergonomics of human-system interaction. Part 910: ISO/CD 9241-910, Framework for tactile and haptic interaction (draft); part 920: ISO 9241-920, Guidance on tactile and haptic interactions (standard published 2009); part 940: ISO/WD 9241-940, Evaluation of tactile/haptic interactions (working draft). Prepared by ISO TC 159/SC 4/WG9 (2009)
10. Landau, S., Gourgey, K.: Development of a talking tactile tablet. Information Technology and Disabilities 7(2) (2001)
11. Miller, C.: Multimedia statistical diagrams. Undergraduate Thesis, Department of Computer Science, The University of Western Ontario (1996)
12. Mittelbach, F., Goosens, M., Braams, J., Carlisle, D., Rowley, C.: *The L^AT_EX Companion*. In: Tools and Techniques for Computer Typesetting, 2nd edn., vol. 3, Addison-Wesley Publishing Company, Boston (2004)
13. Pierce, J.(ed.) Boyce, J., Conrad, J., Dodge, M., Krieger, S., Millhollon, M., Murray, K., Slack, S.E., Sheresh, B., Sheresh, D., Stinson, C., Viescas, J.: 2007 Microsoft Office System Inside Out. Microsoft Press, Redmond (2007)
14. Poh, S.-P.: Talking diagrams. MSc Thesis, The University of Western Ontario, Published as Report 459, Department of Computer Science, The University of Western Ontario (1995)
15. Power, C., Jürgensen, H.: Accessible presentation of information for people with visual disabilities. Univ. Access Inf. Soc. (to appear, 2009)
16. Price, M.: Office 2007 for Seniors. In: easy steps, Southam, United Kingdom (2009)
17. Raman, T.V.: Audio System for Technical Readings. PhD thesis, Cornell University (1994)
18. Schiewe, M., Köhlmann, W., Nadig, O., Weber, G.: What you feel is what you get: Mapping GUIs on planar tactile displays. In: Stephanidis, C. (ed.) Proceedings UAHCI 2009, Part of HCI International 2009, Part II, San Diego, CA, USA, July 19–24. LNCS, vol. 5615, pp. 564–573. Springer, Heidelberg (2009)
19. Sturm, I., Köhlmann, W., Schiewe, M., Jürgensen, H.: Communicating through gestures without visual feedback. In: Makedon, F., Maglogiannis, I., Baillie, L., Pantziou, G., Karkaletsis, V. (eds.) Proceedings of PETRA 2009, Corfu, Greece, June 9–13. ACM International Conference Proceedings Series, p. 8. ACM Press, New York (2009)
20. Völkel, T., Weber, G., Baumann, U.: Tactile graphics revised: The novel BrailleDis 9000 pin-matrix device with multitouch input. In: Miesenberger, K., Klaus, J., Zagler, W.L., Karshmer, A.I. (eds.) ICCHP 2008. LNCS, vol. 5105, pp. 835–842. Springer, Heidelberg (2008)

Improving Screen Magnification Using the HyperBraille Multiview Windowing Technique

Christiane Taras¹, Michael Raschke¹, Thomas Schlegel¹, Thomas Ertl¹,
Denise Prescher², and Gerhard Weber²

¹ Visualization and Interactive Systems Institute, University of Stuttgart
Universitätsstr. 38, 70569 Stuttgart, Germany
<http://www.vis.uni-stuttgart.de/>

² Human-Computer Interaction Research Group, Technical University of Dresden
Nöthnitzer Str. 46, 01187 Dresden, Germany
<http://www.inf.tu-dresden.de/>

Abstract. Screen magnification is an important means to support visually impaired people when working with computers. Many improvements have been made on appropriate software. But unfortunately, in the last years, those improvements were mainly in realization detail. A number of problems remain, that, to our minds, need conceptual rethinking. In this paper, we present a new concept for magnification software. It uses different views to support the user efficiently in different situations. Thereby, it reveals the possibility to build upon current magnification methods and so retain features of current magnification software. The described views are derived from a concept that was originally developed for a tactile graphics display. We found that both topics, rendering for a tactile graphics display and screen magnification, have very much in common. Initial user feedback confirms the usefulness of the concept.

Keywords: Screen Magnification.

1 Introduction

For several years screen magnifiers have been combined with screen readers to support synchronized graphics and speech output (see for example ZoomText and Baum Galileo). Unfortunately, the graphics rendering itself has not been based on information from the screen reader but mainly on enlarging the original screen output or modifications of color schemes. The screen reader information is used only for some small additions like text highlighting or emphasizing the mouse pointer (see for example Freedom Scientific MAGic and Dolphin SuperNova). This has some important disadvantages. One problem is the view becomes very grainy as the pixels of the original screen image are enlarged to squares. Some effort has been made to realize a smoother text and icon rendering [1]. But actually, the results are not totally satisfying. Another problem is that only parts of the desktop content are visible on the screen. Therefore considerable scrolling is necessary to grasp the whole information even when just reading some text.

In this paper we present a new approach for realizing a screen magnifier. This new approach is based on the ideas of the HyperBraille project. The main challenge of this project is to fit the contents of graphical desktop applications to a tactile graphics display with just 120×60 pins. This is quite similar to the task, screen magnifiers are intending to solve. As the architecture of the HyperBraille software is designed in a very flexible way, it can easily be adapted to provide data to a screen magnifier instead of a tactile graphics display. So we managed to combine screen reader information and the idea of magnification in a new way to overcome the above mentioned problems and at the same time sustain the positive features of current screen magnifiers.

2 The HyperBraille Software

The HyperBraille project aims to develop a new tactile graphics display and appropriate software, called HyperReader, that supports visually disabled people in working with modern graphical user interfaces better than ordinary screen readers with speech and Braille output.

To meet this challenge a presentation strategy was defined that renders the contents of graphical applications in multiple views to different regions of the tactile display [3]. The regions defined are header, view list, structure, window titles, detail and body region. The header region, for example, always shows the title of the window the user currently works with or the path of the menu that is currently expanded. So no scrolling is necessary to easily check if one works with the right document or looks for a menu item really in the right menu. The detail region shows additional information on focused objects like the tooltip or information about a comment in a word document. Information about a documents structure like the position of headings or writing mistakes are presented in the structure region. The body region is the main working area. There the contents can be presented in different views that combine tactile graphics and Braille in varying degrees. These views are operating, symbol, layout and original view. Thereby, the operating view is designed for an optimized workflow and presents necessary elements of the user interface in Braille without the need for horizontal scrolling. The layout view provides the user with outlines of the objects on the screen and provides a quick overview of complex objects. The symbol view represents a compromise between operating and the layout view. It tries to preserve the layout as far as possible while using as little space as possible and presenting textual elements in Braille. Furthermore, it presents the user interface elements with special tactile widgets to facilitate a fast recognition of the elements' types. The original view, finally, presents the original screen graphics converted to a black-and-white image by using different image processing algorithms. The body region can be split up into two areas providing, for example, two different views to one window at the same time.

To realize this presentation strategy a graphical screen reader has been developed that builds up an object-based off-screen-model of all applications presented on the screen [2]. Among other things, this representation contains information

about the objects' types, their graphical representations, their positions on the screen and the hierarchical composition of the graphical user interfaces. With these information the screen contents can be recreated completely in whatever representation is appropriate.

HyperReader also provides adaptable rendering and interaction handling that enables the user to explore and control the presented applications. The user can scroll, pan and zoom the views in appropriate ways. The views are synchronized if possible and can easily be switched. The different regions can be shown and hidden as needed. Further, audio output is provided and synchronized with the user's interaction. The integrated configuration management enables the user to influence different components of the HyperReader in various ways easily.

3 Implementation of the Screen Magnifier

Like almost every component of the HyperReader, its rendering pipeline is designed very flexible. Although it was developed to support a tactile rendering, it creates the output by using the visual widget toolkit `Windows.Forms` which is part of the .Net Framework. The components of the off-screen-model are transformed into widgets derived from `Windows.Forms.Control` and composed into panels which are finally rendered into the different regions. The result is provided to the tactile graphics display as bitmap.

For the screen magnifier we have skipped the creation of the output bitmap but use the widgets directly as part of the magnifier window. By that, default mouse and keyboard interaction handlers of `Windows.Forms.Control` can easily be connected to the HyperReader interaction handling, like switching the view, or to special interaction for the magnifier, like enlarging the font size or changing the foreground and background color. Further, predefined controls of the .Net framework with their layout and rendering algorithms can be used. Those also support ordinary and smooth text rendering for every font style and size. So the problem of grainy text rendering can be solved completely.

The magnifier window may be moved and resized freely. This is one possibility how a number of magnifier modes provided by current screen magnifiers are supported. The other way is to integrate the different modes into the different views of our magnifier, mainly the original view, and combine this with the split mode of the body region. The magnifier window can also be placed on a second screen to view original screen output and magnifier output next to each other.

All regions of the HyperReader are also present in the magnifier window. The header, view list, structure, window titles and detail region are mainly text oriented. They are composed of labels that can be surrounded by a border or highlighted. The colors and font type used to render these labels can be changed at runtime via the magnifier menu. The font size can also be changed via pressing accelerators keys `control+plus` or `control+minus`, as it is known from web browsers. The chosen settings for font and color can be saved to the configuration and reloaded at the next start of a session with the magnifier. In the body region the different HyperReader views are rendered and reinterpreted

for the magnifier. Thereby the operating view and the symbol view use the same font settings as the other regions. The layout view uses the configured font type and color but adapts the font size to the current magnification scale. Only in the original view the original text rendering is shown. The magnification scale of the layout view and the original view is independent from the chosen font size but can be synchronized.

In the symbol view markers for different types of user interface elements are used just like for the tactile rendering. But as in visual rendering graphical and textual output can be more easily distinguished than on the tactile display, there are much more design options available. Also, colors can be used. The symbol view tries to keep the original layout as far as possible. But to save space white space is not enlarged with the font size.

The original view is quite similar to the views other screen magnifiers provide. It is based on an enlarged view of the original screen output. But, using the information from the off-screen-model, it is enhanced a lot. The text can be re-rendered smoothly as overlay and also markers as defined for the symbol view can be used to better visualize the types of the shown user interface elements.

All over the magnifier window an enlarged mouse cursor is shown that can be emphasized with a special marker like in other screen magnifiers. Further, a tooltip using the chosen font settings shows additional information on the control elements of the magnifier window like the labels in the view list region.

4 Results

Figure 1 shows screenshots of the current prototype of our screen magnifier. The screenshots are in German as the current prototype of the HyperReader mainly supports German.

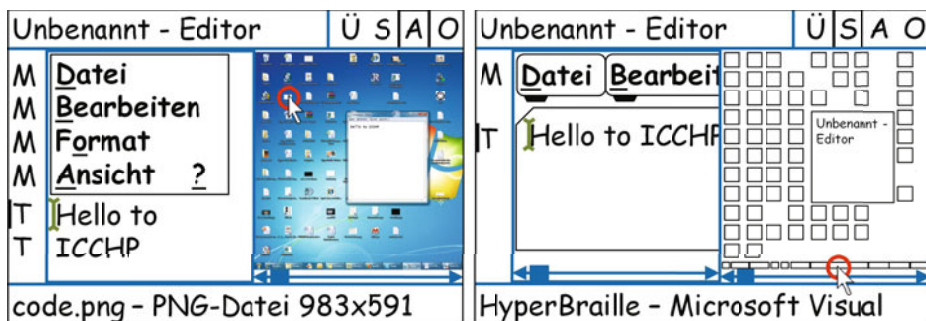


Fig. 1. Two screenshots of the first prototype of the magnifier. In both screens show the header, the view list, the structure, the detail, and the body region which is split. Different views are shown. From left to right there is operating view of the Notepad window, original view of the desktop, symbol view of the Notepad window, and layout view of the desktop.

The left screenshot shows the operating view of the Notepad window at the left of the body region and the original view of the desktop at the right. The right screenshot shows the symbol view of the Notepad window at the left of the body region and the layout view of the desktop at the right. The font size of the operating and the symbol view is comparable to a magnification of the menu and text in the Notepad window by factor 9.

In both screenshots all regions except the window title list are shown. The header region shows the title of the current window "Unbenannt Editor". The view list marks the current view of the active part of the body region (which is left in both screenshots). The detail region shows information on the object under the mouse cursor. The structure region marks the menu items and the text elements of the active part of the body region, as well as the cursor position. Further, a magnified mouse cursor and text caret are shown.



Fig. 2. Screen designs that visualize ideas on improvements for the magnifier

To get some initial user feedback, we have presented our prototype to three visually impaired users. Their judgement showed us that it is worth working further on our concept. They found the split view useful to work with two different documents in parallel. The combination of structure region and operating view proved to be helpful for fast document reading. A negative aspect that was mentioned was that the scrolling effort to reach for example toolbars or the statusbar was not reduced. This had already come to our minds during the implementation. Furthermore, we recognized that the view list region is not necessary for the magnifier, as of course, the users can easily distinguish between the different views. In contrast, for the tactile output, the view list is an essential hint to tell the user, how to interpret the output image. Another difference to the tactile output is that the layout view seems not to be that useful when presented as shown in figure 1. But it could be useful when combined with the original view as optional emphasis of element borders.

Building on this initial feedback we have developed further ideas to improve the magnification. The most important thing is to distinguish between different working areas, like document, toolbars, menu and statusbar. We decided to use the body region in operating view only for presenting the document. Toolbars, menu and statusbar should be available via the header region.

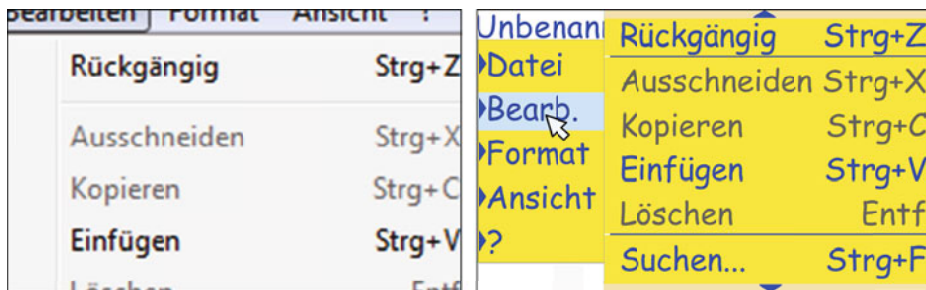


Fig. 3. Visualization of the effect of an optimized menu re-rendering (right) in contrast to an ordinary magnification of the same menu (left) at 10 times magnification. The re-rendering uses text shortening and relocation of symbols to reduce the space needed for the menu. Furthermore, ordinary menu layouting features are utilized that fit the submenu to the screen height. So, the user can easily scroll through the submenu entries without moving the main menu out of the view.

Figure 2 shows screen designs to realize the improvements. The right image shows a design to support document creation with word. Important parts of the ribbon are shown in the header region. Therefore, the header was split in two different scrollable parts. To realize such a view, the application's elements have to be filtered differently for different tasks. An appropriate concept has already been developed for the tactile output of the HyperReader and can be adapted to the magnification mode.

The image also shows, that we have decided to not restrict the different views to the body region any more, but also provide them for the header view. So the user can decide whether to view for example toolbars in pure text view, in symbol view or in original view.

In the left image of figure 2 a default rendering for the menubar is shown. It is opened either by pressing the Alt key or by a mouse click on the header region. It should be rendered as overlay over the whole magnifier window but be bound to the screen size. So for sub menus features of ordinary menu rendering can be used, that layout menus to fit into the screen size. This further reduces scrolling effort and improves overview as figure 3 visualizes. The left image of figure 3 shows an ordinary magnification of the menu using the same scaling factor as the right image represents.

There, also other ways to reduce the space that is necessary to render a menu are shown. One method is to shorten known texts as in the image "Bearbeiten" by "Bearb.". This concept has already been implemented for the tactile symbol view. For a number of applications this could be also realized in current screen magnification software by providing adapted localization files. Further, we have placed the symbol that indicates the presence of a submenu left of the menu. Normally, it is placed right and symbols like check marks for checkable menus are placed left. So, appropriate space is needed left and right of a menu's text, even if no symbols are shown for this menu. With the relocation of the submenu symbol we save half of this empty space. Modifications like this are only possible

with re-rendering and appropriate knowledge about the element that should be rendered.

5 Conclusion and Future Work

We have presented ongoing work on developing a new approach to screen magnification based on the HyperReader multiview windowing technique. The new screen magnifier solves some major problems of currently available ones but also preserves their advantages.

Of course, implementing accurate re-rendering for each and every application is a huge effort and it depends heavily on the accessibility of the applications. But as our new concept combines different views, the support for different applications can be introduced step by step while all the features of current screen magnification software could be provided by the original view of poorly accessible applications.

In the future, we want to improve our current prototype according to the initial user feedback and the designs shown above. Furthermore, we plan to examine if some of the improvements of the magnifier can also be adapted to the tactile output.

Acknowledgments. The HyperBraille project is sponsored by the German Ministry of Economy and Technology (BMWi) under the grant number 01MT07004. Only the authors of this paper are responsible for its content.

References

1. Blenkhorn, P., Evans, D.G., Baude, A.: Full-Screen Magnification for Windows Using DirectX Overlays. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 10, 225–231 (2002)
2. Kraus, M., Völkel, T., Weber, G.: An Off-Screen Model for Tactile Graphical User Interfaces. In: Miesenberger, K., Klaus, J., Zagler, W.L., Karshmer, A.I. (eds.) *IC-CHP 2008*. LNCS, vol. 5105, pp. 865–872. Springer, Heidelberg (2008)
3. Schiewe, M., Köhlmann, W., Nadig, O., Weber, G.: What you feel is what you get: Mapping guis on planar tactile displays. In: Stephanidis, C. (ed.) *HCI (6)*. LNCS, pp. 564–573. Springer, Heidelberg (2009)

Three-Dimensional Tactile Models for Blind People and Recognition of 3D Objects by Touch: Introduction to the Special Thematic Session

Yoshinori Teshima

Advanced Manufacturing Research Institute,
National Institute of Advanced Industrial Science and Technology,
1-2 Namiki, Tsukuba, Ibaraki 305-8564 Japan
yoshinori.teshima@aist.go.jp

Abstract. Blind people can recognize three-dimensional shapes through tactile sensations. Therefore, effective models are useful in tactile learning. In our special thematic session, we will focus upon three-dimensional tactile models which developed by using the most advanced technologies. Another focus of the special thematic session is devoted to the recognition mechanisms of 3D objects based on tactile information. This includes the problem to recognize the shapes of 3D objects from 2D tactile plane figures.

Keywords: tactile sensation, tactile learning, tactile teaching, teaching materials, 3D tactile models in science, 3D tactile models in mathematics, 3D tactile models in art, recognition mechanisms of 3D objects, tactile plane figures.

1 Introduction

In the last few decades, remarkable progress has been made in the development of exact 3D solid models. This is mainly attributable to modern technology and techniques such as information technology, shape measurement, and digital manufacturing. Exact models are useful as tactile teaching materials. Such models enable blind people to experience the world through tactile sensations. They are also instructive for sighted people. Some simple improvements can make these models more helpful for blind people.

One purpose of this special thematic session (STS) is to bring together experts from around the world to present and discuss state-of-the-art research and development activities and future perspectives in three-dimensional tactile models.

Another purpose of this STS is to discuss the recognition mechanism of shapes of 3D objects by blind and visually handicapped people. Recognition mechanisms of 3D objects based on tactile information differ from those based on visual information.

Does there exist a systematic method for the tactile observation of 3D tactile models so that blind students can recognize them correctly and efficiently? Blind people can recognize the shapes of 3D objects by touching them; however, it is quite difficult to recognize the shapes of 3D objects from 2D tactile plane figures. This STS intends

to discuss solutions to these problems that currently hamper the education of blind and visually handicapped students.

3D models are very old but the newest topics. There are inexhaustible possibilities for the developments of tactile 3D models in various subjects. Such models are significant for sighted people as well. That is, not only visual information but also use of tactile sensation will bring a better recognition of 3D objects to sighted people. I hope this STS plays a important role in initiating further study in this direction. This STS is a first step toward a faraway journey.

I encourage you to participate in our STS and discuss current problems with us. Your frank and open remarks will enliven our STS. Welcome to our newly started STS in ICCHP 2010.

Models of Mathematically Defined Curved Surfaces for Tactile Learning

Yoshinori Teshima¹, Tohru Ogawa^{1,2}, Mamoru Fujiyoshi³, Yuji Ikegami^{1,4},
Takeshi Kaneko⁵, Susumu Oouchi⁵, Yasunari Watanabe^{1,4}, and Kenji Yamazawa⁴

¹ Advanced Manufacturing Research Institute, National Institute of Advanced Industrial
Science and Technology, 1-2 Namiki, Tsukuba, Ibaraki 305-8564 Japan
yoshinori.teshima@aist.go.jp, ogawa-t@koalanet.ne.jp,
yikegami@riken.jp, wistajp@yahoo.co.jp

² University of Tsukuba (emeritus professor), Tennodai, Tsukuba, Ibaraki 305-8577 Japan
ogawa-t@koalanet.ne.jp

³ Organization of the Study for College Admissions, National Center for University
Entrance Examinations, 2-19-23 Komaba, Meguro, Tokyo 153-8501 Japan
fujiyosi@rd.dnc.ac.jp

⁴ Rapid Engineering Team, RIKEN Advanced Science Institute(ASI), 2-1 Hirosawa,
Wako, Saitama 351-0198 Japan
yikegami@riken.jp, wistajp@yahoo.co.jp, kyama@riken.jp

⁵ National Institute of Special Needs Education, 5-1-1 Nobi, Yokosuka, Kanagawa
239-8585 Japan
kaneko@nise.go.jp, oouchi@nise.go.jp

Abstract. Several accurate models of mathematically defined curved surfaces were constructed for use in tactile learning. Exact shape data were generated on a personal computer (PC) using mathematical or computer-aided design (CAD) software. Then tactile models were constructed by layered manufacturing, which is well suited for curved surfaces. This method is flexible in that the equation parameters and model scale can be changed easily. A recognition test performed on several models showed their potential usefulness for tactile learning.

Keywords: tactile 3D model, mathematically defined curved surface, layered manufacturing.

1 Introduction

In the past, there has been no systematic study on the development of three-dimensional (3D) mathematical models that can be used as tactile learning materials for the blind. A famous example is that of renowned German mathematician Felix Klein, who started the development of 3D mathematical models for visual observation around 1870 [1]. This required a host of the best mathematicians to work in collaboration with skilled workmen. They developed many wonderful models, but since production ceased, these models are rarely found these days and, if found, have



Fig. 1. Models produced in Germany around 1870 under a project led by Klein (plaster models)



Fig. 2. Model 1: Dupin cyclide constructed by stereolithography; width = 7 cm

to be treated with utmost care. Fig. 1 shows a photograph of some of Klein's models encased in a glass showcase at the mathematics department of the University of Tokyo.

The situation is the same in almost all museums. There are many objects in museums all over the world; however, these objects are often kept in glass showcases, and blind people cannot touch or perceive them. This problem hampers the learning experience of the blind and visually handicapped.

We have been involved with the 'Research on the recognition of 3D objects by visually handicapped persons and development of 3D geometrical teaching materials' project in Japan since 2006 [2] [3]. One of our main aims is to develop teaching materials to enrich the knowledge of the blind. Many models have been developed thus far.

Times have changed. The situation today is quite different from that in 1870. Now, we can construct mathematical models without the assistance of the best mathematicians and skilled workmen. Model 1 (Fig. 2) is the classic mathematical curved surface, the Dupin cyclide, which is also shown in Fig. 1.

For the blind, many shapes are beyond their ability to perceive with touch, either due to their gigantic or microscopic size or due to their abstract nature, for example, mathematical models.

In this paper, we describe our 3D models of mathematically defined curved surfaces, which can be used for tactile learning, as indicated by the results of a recognition test.

2 Development of Models

In this study, tactile learning materials were developed using modern technology and techniques in order to enrich the knowledge of the blind. The exact shape data were obtained from the equations of the surfaces using a personal computer (PC) running mathematical software such as Mathematica and Maple or a computer-aided design (CAD) suite. The obtained shape data file was converted into the STL format; this format is native to stereolithography, and is also available for other specific processes in layered manufacturing. And finally, the exact model was constructed using a layered manufacturing machine.

Layered manufacturing, better known as rapid prototyping, allows automatic construction of 3D objects. Several methods exist, each using a different material. In this

study, we used stereolithography (material: photocurable resin), selective laser sintering (material: nylon powder), and 3D printing (material: plaster powder).

Note that although the ideal thickness for these surfaces is zero, this is not practical and so the models were made sufficiently thick and strong.

3 Models of Mathematically Defined Curved Surfaces

Our main results are outlined in this section. The photographs of models shown in this paper are accompanied by the model name, specific process used in layered manufacturing, and length or height of the model. The two half sections of the models are shown to the right of the photographs of Models 3, 4, 5, 6, 7, 8, and 10.

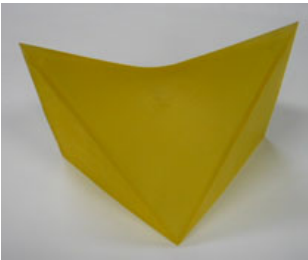


Fig. 3. Model 2: hyperboloid of one sheet constructed by stereolithography; height = 10 cm



Fig. 4. Model 3: Klein bottle, Left: constructed by stereolithography, height = 10 cm, Right: constructed by 3D printing, height = 20 cm

Models 1 and 2 (Fig. 2 and 3) are classic mathematical models and, therefore, they are not explained in this paper. The Klein bottle (Fig. 4) is a non-orientable surface. The surface has no distinct 'inner' and 'outer' sides. And the inside space of the bottle is linked to its outside space. But we cannot create such a model in 3D because a self-intersection is unavoidable. The self-intersection is avoidable in 4D. Model 3 is an exact model of the Klein bottle. We can touch the self-intersection in the pair of equal sections (Fig. 4, right-hand side).



Fig. 5. Model 4: ring torus, constructed by selective laser sintering; $d = 7.5$ cm, $r = 2.5$ cm



Fig. 6. Model 5: ring torus, constructed by selective laser sintering; $d = 5$ cm, $r = 2.5$ cm

Consider an ordinary torus, which is a surface with a hole (Fig. 5 and 6). Let d denote the radius from the centre of the hole $(0,0,0)$ to the centre of the torus tube and r denote the radius of the tube. There are three types of tori depending on the relative values of d and r . The condition $r < d$ corresponds to a ring torus (Fig. 5 and 6); $r = d$, a horn torus, which is tangential to itself at the point $(0,0,0)$ (Fig. 7); and $r > d$, a self-intersecting spindle torus (Fig. 8 and 9).



Fig. 7. Model 6: horn torus, constructed by selective laser sintering; $d = 2.5$ cm, $r = 2.5$ cm

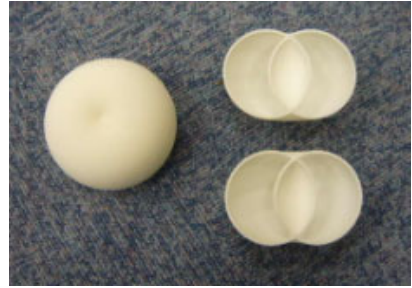


Fig. 8. Model 7: spindle torus, constructed by selective laser sintering; $d = 1.25$ cm, $r = 2.5$ cm



Fig. 9. Model 8: spindle torus, constructed by selective laser sintering; $d = 0.625$ cm, $r = 2.5$ cm

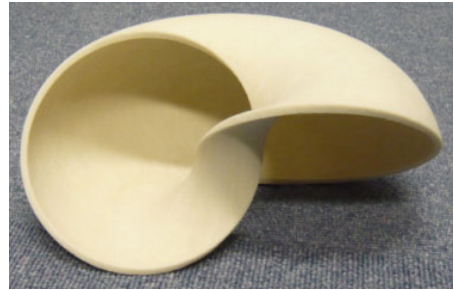


Fig. 10. Model 9: our original math surface constructed by 3D printing; width = 20 cm

The horn torus and spindle torus do not have a hole. The abovementioned tori are regarded as the loci of circular movement of a circle, which is perpendicular to the horizontal plane. If we bend the perpendicular circle backward by 45° , a hole appears at the locus of circular movement despite the relative size $r > d$. If we consider a semicircular movement (180°) of the inclined circle, we obtain an interesting shape (Fig. 10).

Now, we consider circular movement in a vertical plane. Then, the circle always turns its face towards the vertical direction. As a result, we obtain a novel surface (Fig. 11). It should be noted that the half sections (centre of photograph of Fig. 11) are incorrect because the two circles in the cross section were replaced by two disks. This

incorrect model was used in our recognition test. We corrected this error after the recognition test (right-hand-side photograph in Fig. 11).

Another novel model is obtained by considering the elliptical movement of an ellipse (Fig. 12). Models 9, 10, and 11 are original mathematical curved surfaces obtained in this manner [4].



Fig. 11. Model 10: our original math surface constructed by selective laser sintering; Left: complete model, Middle: incorrect partitioned models, Right: corrected partitioned models, height = 6 cm



Fig. 12. Model 11: our original math surface constructed by selective laser sintering; height = 6cm



Fig. 13. A ring is subsidiary tool. Students move the ring and imagine the locus of it. Another subsidiary tool is a stick (Fig. 14, 15, and 16) which is the rotational axis of circular movement.

4 Recognition Test for Models

A recognition test was performed on some of the models. Nine blind high-school students were asked to fill up a questionnaire with 23 questions. The target models were Models 4, 5, 6, 8, and 10—all of which can be regarded as the locus of circular movement of a circle. Subsidiary tools shown in Figs. 13, 14, 15, and 16 were also used in the recognition test.

Before answering the questionnaire, each student was required to hold a ring (Fig. 13) and rotate it by 360° without changing his/her standing position. All students were able to imagine the locus (Models 4 and 5) of the ring. The configuration of the ring and the rotational axis was then explained to them using the subsidiary tool shown in Fig. 14.

Each student had to imagine the locus (Models 6 and 8) of circular movement of the ring in Figs. 15 and 16. The students had to then determine the shape of the cross

sections before and after tactile observation of the equal halves. They also had to determine the position of the rotational axis in the complete models and the partitioned models.



Fig. 14. Subsidiary tools to understand ring torus



Fig. 15. Subsidiary tools to understand horn torus

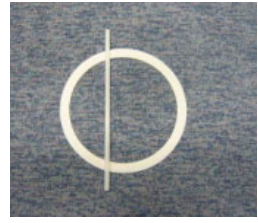


Fig. 16. Subsidiary tools to understand spindle torus

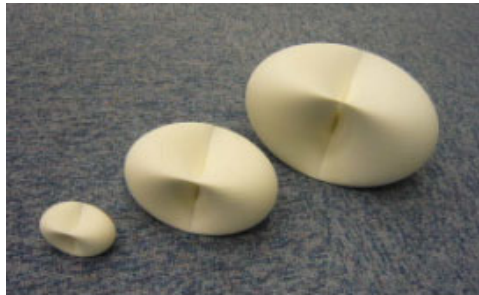


Fig. 17. Different sizes of Model 10. Left: height = 3 cm, Middle: height = 6 cm, Right: height = 9 cm

The percentage of correct answers given by the students was very high (89% or 100%) for each question on Models 4, 5, and 6. However, the percentage of correct answers for the question on the shape of the cross section of Model 8 was very low (22%). Most of the students answered that 'there is a small circle in the ellipse'.

The tester then made the students hold the ring such that it was horizontal and move it along a circle in the vertical plane. As a result, all the students could accept the shape of Model 10.

A majority (67%) of students selected the medium-sized model shown in Fig. 17 as the appropriate size for tactile observation.

Finally, the students were asked to freely voice their opinions about the test. Almost all answered that the models were very useful in understanding 3D shapes.

5 Discussion

Various 3D models, explained in section 3, were constructed in this study by layered manufacturing on the basis of 3D shape data generated using mathematical or CAD

software. This combination of methods is excellent and has several great advantages for the development of tactile learning materials in mathematics.

Models of various mathematical curved surfaces were created by changing the equation parameters; Models 4, 5, 6, 7, and 8 are a few examples. Further, the layered manufacturing technique allowed the model scale (Fig. 17) to be easily changed.

Models 4, 5, 6, and 8 were used in the recognition test. Students found the cross-sectional shape of Model 8 difficult to understand (Fig. 9). This is similar to the findings of an interesting study in psychology [5]. The study investigated the tactile perception of the compound figures by the blind. A typical example of the type of compound figure is shown in Fig. 18 (i). The figure can be recognized in two ways: as two overlapping circles (Fig. 18 (ii)) or three small divisions (Fig. 18 (iii)). The paper pointed out that some blind subjects recognized the figure as Fig. 18 (iii). A set of figures in Fig. 19 was applied to a test for the recognition shift from Fig. 18 (iii) to Fig. 18 (ii). But it was difficult for such subjects to recognize the figure as Fig. 18 (ii).

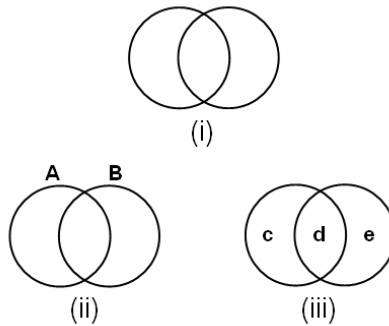


Fig. 18. Recognition of two overlapping circles

Though the study [5] treated 2D figures, the findings are closely related to the recognition of cross sections of each torus in the present recognition test. Fig. 20 is a set of cross sections from our recognition test. Our configurations (Fig. 20) include two contact circles (cross section of Model 6) but Fig. 19 does not include them. They are indispensable for systematic recognition because the two contact circles are a transition configuration between two non-touch circles (Fig. 19 (1)) and two overlapping circles (Fig. 19 (2)).

We suppose that the students will be able to understand the cross-sectional shape of Model 8 more easily if they are given Model 7 (Fig. 8) during the recognition test. Model 7 compensates for the great parameter difference between Models 6 and 8. Further, the shape of the overlapped part in the cross section of Model 7 is not similar to a circle, which eliminates the possibility of students getting confused like in the case of Model 8 ('there is a small circle in the ellipse'). This will be confirmed in future recognition tests.

All models described in this paper can be used by sighted people as well. There are many possibilities for the development of tactile 3D models of mathematically defined curved surfaces. We hope that this study plays a major role in initiating greater research in this direction.

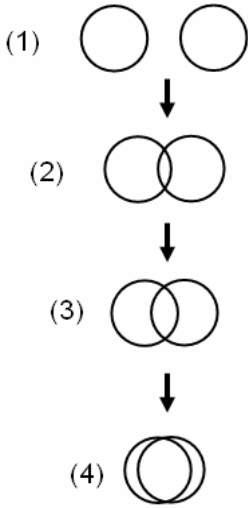


Fig. 19. A set of figures for recognition shift in [5]

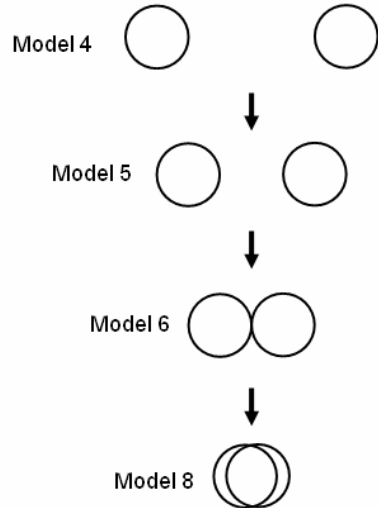


Fig. 20. A set of cross sections in the present recognition test

Acknowledgments. This study was partially supported by a Grant-in-Aid for Scientific Research (A) (18200049) from the Japan Society for the Promotion of Science (JSPS).

References

1. Hilbert, D., Cohn-Vossen, S.: *Geometry and the Imagination*. Chelsea Publishing Company, New York (1952) (English edition); Original edition: *Anschauliche Geometrie*. Springer, Heidelberg (1932)
2. Teshima, Y., Ikegami, Y., Yamazawa, K., Watanabe, Y., Fujiyoshi, M., Kaneko, T., Oouchi, S.: Geometrical teaching materials for visually handicapped person: they see forms by touching. *The Journal of the International Society for the Interdisciplinary Study of Symmetry (ISIS-Symmetry)* 2-4, 464–467 (2007)
3. Teshima, Y., Fujiyoshi, M., Ikegami, Y., Kaneko, T., Matsuoka, A., Nakano, T., Oouchi, S., Tanaka, A., Watanabe, Y., Yamazawa, K.: Development of three-dimensional geometrical models for tactile teaching materials. In: *Proceedings of Tactile Graphics 2008 The fourth international conference and exhibition on tactile diagrams, maps and pictures*, Birmingham, UK, p. 30 (2008)
4. Teshima, Y., Ogawa, T.: Loci of circular movement of circle and their layered manufacturing models. *The Journal of the International Society for the Interdisciplinary Study of Symmetry (ISIS-Symmetry)* 1-4, 244–247 (2009)
5. Kimura, M.: Perception of the compound figures by tactile motor scanning. *Japanese Journal of Psychology* 43(1), 1–12 (1972)

Enlarged Skeleton Models of Plankton for Tactile Teaching

Yoshinori Teshima¹, Atsushi Matsuoka², Mamoru Fujiyoshi³, Yuji Ikegami^{1,4},
Takeshi Kaneko⁵, Susumu Oouchi⁵, Yasunari Watanabe^{1,4}, and Kenji Yamazawa⁴

¹ Advanced Manufacturing Research Institute, National Institute of Advanced Industrial
Science and Technology, 1-2 Namiki, Tsukuba, Ibaraki 305-8564 Japan
yoshinori.teshima@aist.go.jp, yikegami@riken.jp,
wistajp@yahoo.co.jp

² Department of Geology, Niigata University, 8050 Ikarashi 2-no-cho, Nishi-ku, Niigata
950-2181 Japan
matsuoka@geo.sc.niigata-u.ac.jp

³ Organization of the Study for College Admissions, National Center for University Entrance
Examinations, 2-19-23 Komaba, Meguro, Tokyo 153-8501 Japan
fujiyosi@rd.dnc.ac.jp

⁴ Rapid Engineering Team, RIKEN Advanced Science Institute(ASI), 2-1 Hirosawa,
Wako, Saitama 351-0198 Japan
{Yikegami, kyama}@riken.jp, wistajp@yahoo.co.jp

⁵ National Institute of Special Needs Education, 5-1-1 Nobi, Yokosuka, Kanagawa
239-8585 Japan
{Kaneko, oouchi}@nise.go.jp

Abstract. Blind people can learn about real objects through tactile sensations. However, the objects are sometimes rather small to observe by touch. Enlarged three-dimensional (3D) models for such tiny objects are useful for tactile teaching. This paper presents the first exact models of radiolaria and foraminifera. Their 3D shape data are measured using micro X-ray CT, and their exact 3D models are constructed using layered manufacturing. A recognition test showed the immense usefulness of these enlarged models for blind students in learning the shapes of micro-organisms.

Keywords: microorganism, plankton, radiolarian, foraminifera, enlarged model, 3D model, X-ray CT, layered manufacturing, tactile teaching, recognition test.

1 Introduction

Direct observation of real objects is very important in science class. The importance further increases in the case of visually impaired (VI) students. Biology text books have various photos and diagrams of creatures, but the VI students do not find it easy to understand such two-dimensional (2D) graphics. These students can learn the shapes of real creatures through tactile sensations. However, real creatures are sometimes unsuitable for tactile observation because some of them are too big, small, soft, or dangerous for the same. Three-dimensional (3D) models and tactile diagrams are useful for learning the shapes of such creatures.

It is often difficult for the VI students to understand 2D tactile diagrams without any explanation. Before using them, the students have to acquire an ability that enables them to understand correspondence between 2D diagrams and 3D shapes [1]. When the teacher offers the students three kinds of tactile materials, the following order is desirable: the real thing, 3D model, and 2D tactile diagram [2]. Effective usage of the three kinds of tactile materials will bring the VI students a good understanding of the natural world.

Models of microscopic objects are the main focus of this study. We focus on the following two types of micro-organisms: radiolarians and foraminifers. Radiolarians are one of the major groups of marine protistan plankton with silica (SiO_2) skeletons. Foraminifers are another protistan group with calcium carbonate skeletons living as benthos and plankton in a wide variety of marine environments. These two groups appeared in the Cambrian time, approximately 500 million years ago, and are still alive in the modern oceans. They produce sophisticated skeletons with highly diversified morphology. Radiolarian and foraminiferal fossils are utilized as geochronological and paleoenvironmental indicators [3], [4], [5].

Because radiolarians and foraminifers are less than 1 mm in size, it is not possible to recognize their shapes by touching them. We have to use a microscope to obtain their enlarged images, which, however, gives only 2D information, and is of no use to the blind.

There are some 3D models of radiolarians and foraminifers in museums [6]. However, because these were handmade models whose shapes are based on 2D information of SEM (Scanning Electron Microscope), such models do not show the correct shapes images.

Models that will be described in this study have exact shapes. A 3D shape measurement offers their exact shape data, and an automatic manufacturing creates their exact 3D models. These are the first exact models of radiolarians and foraminifers.

2 Development of Models

In the last few decades, remarkable progress has been made in the measurement of 3D shapes. One such technology is the X-ray computed tomography (CT). Recently, X-ray CT has been employed for studying even small objects.

The shape of the skeletons of radiolaria and foraminifera was determined using X-ray CT. The shape data file was then converted into the stereolithography (STL) format, which is the file format used in the layered manufacturing technique. Finally, exact models of radiolaria and foraminifera were constructed using a layered manufacturing machine.

A high-resolution micro X-ray CT SkyScan1172 which includes an 11 Megapixel X-ray digital camera was utilized in this study. The voltage and current of X-ray source were 60 kV and 68 μA . The spatial resolution of scan was 0.77 μm . The reconstruction angular range and the rotation step of a sample were 187.8° and 0.3°, therefore 626 projections were obtained.

There are several techniques involved in layered manufacturing. In this study, selective laser sintering (material: nylon powder) and 3D printing (material: plaster powder) were used to construct the models.

3 Models of Radiolaria and Foraminifera

Our models are shown in this section. These are the first exact models of radiolaria and foraminifera. Therefore, scientists researching radiolaria and foraminifera had long been looking forward to emergence of these models.

We also constructed partition models of radiolaria (tower type) and foraminifera. Partition models are very effective in order to observe their complex inner structures.

4 Recognition Test for Models

A recognition test was performed on our models. Ten completely blind high-school students were asked to fill up a questionnaire, which contained 13 questions. Several questions revolved around the shape of tactile models. The following are excerpted questions and their percentages of correct answers.

The radiolarian in Fig. 1 has six dividers inside. First, the students counted the number of the dividers using the completed model (Fig. 1, left). The percentage of correct answers obtained was 40%. Second, they counted the number of dividers using the partitioned model (Fig. 1, right), and the percentage of correct answers increased to 90%.



Fig. 1. Radiolarian (Tower type), Scientific name: *Svinitzium pseudopuga* Dumitrica, Real size: 200 μm , Shape data: Micro X-ray CT, Left (complete model): 24 cm (ratio: 1200 times), constructed using selective laser sintering (nylon powder), Right (partitioned model): 18 cm (ratio: 900 times), constructed by 3D printing (plaster powder)

The three models shown in Fig. 2 are based on the same shape data. One of the questions was regarding the appropriate size that could be observed of the whole shape in the figure. The choices were between the centre (21 cm) and the right (33 cm), and 90% of the students chose the smaller model. Another question was about the tactile feel of the models. The choices were between the left (plaster) and the centre (nylon) models, and 90% of the students chose the nylon model.

The two models (top left: 10 cm and right: 20 cm) in Fig. 3 are based on the same shape data. One of the questions was about the appropriate size that could be observed of the whole shape in the figure. All the students preferred the smaller model. In another question, the students counted the number of small rooms inside the foraminifera-60% of the students chose the correct answer (nineteen rooms), and 20% chose incorrect but nearly correct answers (eighteen and twenty rooms).



Fig. 2. Radiolarian (Two horn type)

Scientific name: *Pantanelium riedeli* Pessagno, Real size: 300 μm , Shape data: Micro X-ray CT, Left: 24 cm (ratio: 800 times), constructed using 3D printing (plaster powder), Centre: 21 cm (ratio: 700 times), constructed using selective laser sintering (nylon powder), Right: 33 cm (ratio: 1100 times), constructed by selective laser sintering (nylon powder)

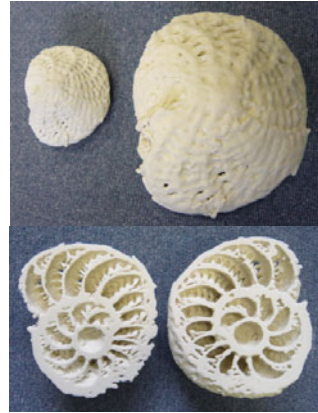


Fig. 3. Foraminifera

Scientific name: *Elphidium*, Real size: 500 μm , Shape data: Micro X-ray CT Constructed using 3D printing (plaster powder), Top Left: 10 cm (ratio: 200 times), Top Right (complete model) & bottom (partitioned model): 20 cm (ratio: 400 times)

In the last part of the questionnaire, the students were asked to freely write down their opinions regarding the test. Almost all the students opined that the models were very useful for understanding the shapes of radiolaria and foraminifera.

Acknowledgments. We are grateful to TOYO Corporation for providing the micro X-ray CT data (SkyScan1172). This study was partially supported by a Grant-in-Aid for Scientific Research (A) (18200049) from the Japan Society for the Promotion of Science (JSPS).

References

1. Toriyama, Y.: Tactile Graphics for teaching biology. In: Proceedings of Tactile Graphics 2008, Birmingham, UK, p. 30 (2008)
2. Takei, Y.: Learning Invertebrate animals in junior high school for the blind: JASEB News Letter, Bulletin of Japanese Association of Science Education for the Blind (27), 6–9 (2008) (in Japanese)
3. Matsuoka, A.: Living radiolarian feeding mechanisms: new light on past marine ecosystems. *Swiss Journal of Geoscience* 100, 273–279 (2007)
4. Matsuoka, A.: Current activities of radiolarian research. *Forma* 14(1-2), 199–204 (1999)
5. Tappan, H., Loeblich, A.R.: Foraminiferal evolution, diversification and extinction. *Journal of Paleontology* 62, 695–714 (1988)
6. Geological Museum Web, <http://www.gsj.jp/Muse/eng/index.html>

Reproduction of Tactile Paintings for Visual Impairments Utilized Three-Dimensional Modeling System and the Effect of Difference in the Painting Size on Tactile Perception

Susumu Oouchi¹, Kenji Yamazawa², and Lorreta Secchi³

¹ National Institute of Special Needs Education, Japan
5-1-1 Nobi, Yokosuka-shi, Kanagawa, Japan
ouchi@nise.go.jp

² RIKEN (The Institute of Physical and Chemical Research)
Hirosawa 2-1, Wako-shi, Saitama, Japan
kyama@riken.jp

³ The Francesco Cavazza Institute for the blind, Italy
via Castiglione, 71 40124 Bologna – Italy
loretta@cavazza.it

Abstract. It is difficult for blind persons to appreciate painting. To facilitate the appreciation by tactile perception, the Francesco Cavazza Institute for blind in Italy is developing three-dimensional tactile painting for blind persons. We are developing a system which product tactile paintings utilized three-dimensional model making technology cooperation with Cavazza Institute. This method enables us to manufacture paintings of varied sizes. In this study, we examined the possibility of down-sizing products. We find blind persons to be useful means of materials which reaffirm the painting image.

Keywords: tactile painting, three-dimensional modeling, blind student.

1 Introduction

The principles of reproduction and the way of appreciation for tactile painting for blind persons were developed by Cavazza institute for the blind in Italy. The tactile paintings are made of plaster. This led blind persons to appreciate paintings with using tactile clues. It seems impossible for blind persons to appreciate paintings same as sighted persons, but tactile paintings contributed to get close to paintings for them[7].

The museum of tactile paintings developed a system to render plane paintings into semi-three dimensional versions. This relief technology can make more tangibly represent shapes by two-dimensionally compressing three dimensional objects, and represent the senses of depth and distance with layer structures. This technology has enabled representation of the painting's spatial composition and structural characteristics with the restricted thickness[7].

On the other hand, we recognize a problem that haptic memory is easy to lose compared to visual perception[3][8].

Sighted person can enjoy reproduction painting[2]. If blind person have replica of tactile paintings on hand, they will be able to continue to keep tactile image of the picture after the appreciation of original tactile paintings[1][4][5].

We are developing a system which product tactile paintings utilized three-dimensional model making technology cooperation with Cavazza institute. We scan in three-dimensional digital data of original tactile painting using 3D scanner. From the data, we can product replica of original tactile picture using 3D output device. This method enables us to reproduce replicated paintings of varied sizes[6].

We produce replicated paintings reduced original tactile painting by this method. On this occasion, we have the problem that tactile pictures are appreciated by tactile perception. To represent relief image in detail, it is necessary to free up suitable size of tactile picture for tactile perception.

In this study, we examined the possibility of utilization of these down-sizing products on tactile exploration.

1.1 Three-Dimensional Modeling System

The system summary is as follows:

Plaster semi-three dimensional model of painting reprodced from original painting image. We took the example of the “Mona Lisa” painted by Leonardo da Vinci in this study. The reproduction task entrust specialist of relief sculpture. Consideration points in the adaptation from two-dimensional image to three-dimensional image are follows:

- (1) to adapt three-dimensional space to layered structure.
- (2) to make compressed representation of three-dimensional image of the characters or things.
- (3) to make a appropriate modification based on property of tactile perception.

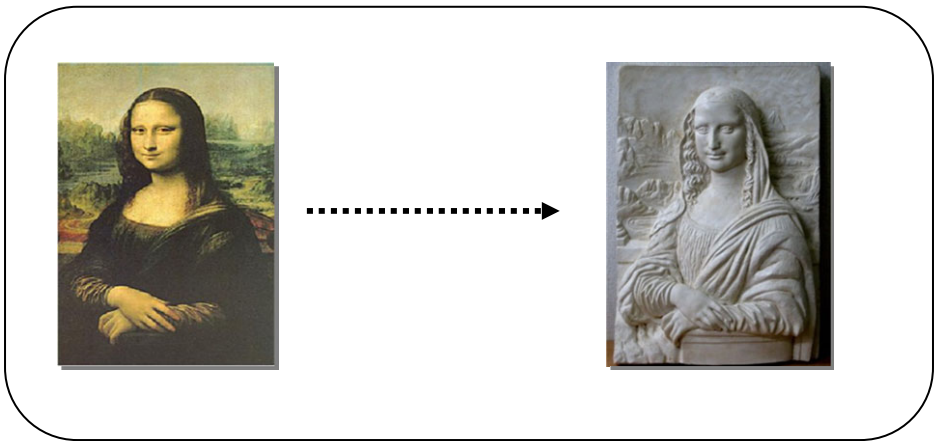


Fig. 1. Reproducing original painting to tactile painting

1.2 Process from Three-Dimensional Digitizing to Modeling

- (1) Acquisition or production of three-dimensional digital data of tactile picture
(Utilized by three-dimensional modeling software or three-dimensional scanner)
- (2) Translated from three-dimensional image data to modeling data through the use of CAM software
- (3) Molding operation by three-dimensional molding machine.

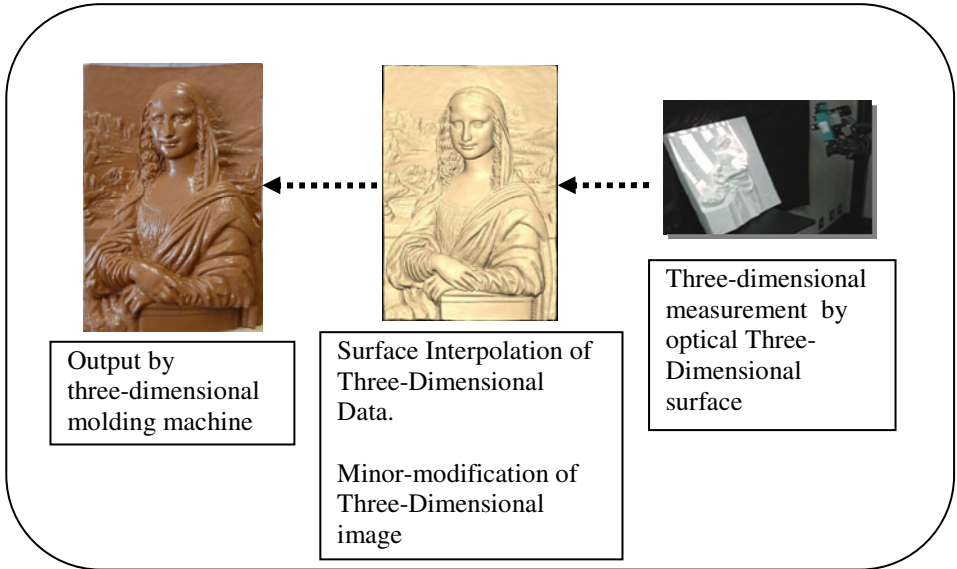


Fig. 2. Process of three-dimensional modeling of tactile painting

2 Verification Experiment

2.1 Participants

Participants were 10 students of high school for the visually impaired(7 men, 3 women). Age average They are all braille reader.

2.2 Materials

(1)Original relief painting “Mona Lisa”

In this experiment, we involved relief painting “Mona Lisa”(fig.1). This material was developed by Anteros Museum in Italy. Relief sculptor adapted the painting by Leonard da Vinci to semi-three dimensional relief. It was 59 cm long, 39 cm wide and 9cm height(max). This relief was made of alabaster.

The Structural Principle of tactile painting

We think that the fundamental principles of the tactile art are the compression and stratification of three-dimensional space, and the transformation of the contents to be

suiting to tactile observation. Especially the compression and stratification are definitely essential to reproduce the cubic effect, sense of depth and perspective.

Compression

Compression is the process in which two-dimensional contents are once transformed to an actual three-dimensional image, and then compressed to a direction to form half solid. For the parts of the art to be emphasized, this compression is a little weakened so that it can provide the sense of depth effectively to tactile observers.

Stratification

Stratification means to reproduce the depth contents in several layers. The depth is interpreted to layers to make it easy for tactile observers to grasp the view perspective.

(2) Three dimensional molding materials

We produced 3 cutdown size materials of “Mona Lisa” using same three dimensional data.

A (about B5size) 15.8cm×23.0cm B (about B7size) 7.9cm×11.5cm C (about B9size) 3.9cm×5.7cm

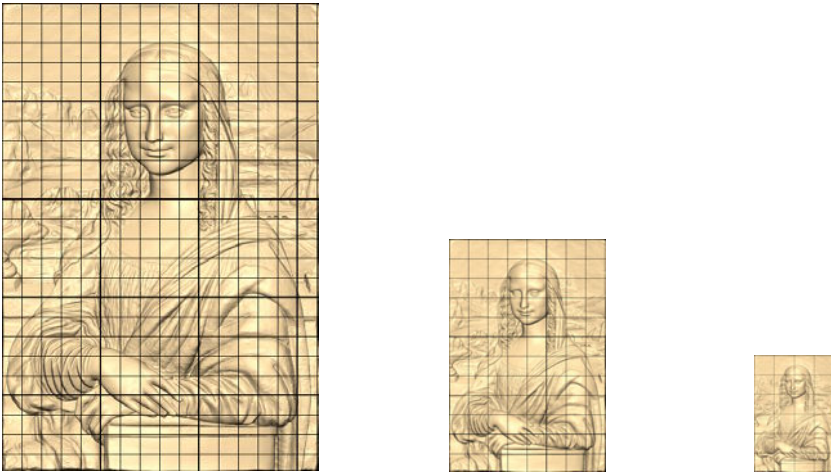


Fig. 3. Cutdown materials produced by molding(B5, B6, B7size)

2.3 Procedure

The experiment was performed individually

At the beginning, Participants appreciated plaster material of relief painting “Mona Lisa”.

In this regard, experimenter expressed in language about contents or color represented in that painting. Participants explored with both hands.

After that, participants appreciated three dimensional molding materials in small size according size A, B, C.

After the appreciation, experimenter asked participant a few questions about recognition degree of every material in conformity with the investigation items. Participants answered questions orally.

2.4 Result

(1)Recognition of plaster material of relief painting “Mona Lisa”

Asked about shape of the material, all participants said the material easily understood and was rich in detail or specifics compared with two-dimensional tactile picture. They thank the material contribute to pictorial appreciation. No one disputed size of the material.

(2) Recognition of reduced materials in size of relief painting “Mona Lisa”(Three dimensional molding materials)

1) Reduction version A (15.8cm×23.0cm)

Asked about shape of the material, 3 participants said “easy to understand”, 7 participants said “to understand”.

2 participants said to be able to understand sufficiently only the material of reduction version A.

8 participants said to be easy to understand.

2)Reduction version B (7.9cm×11.5cm)

3 participants said to be easy to understand, 7participants said to be not easy to understand.

3)Reduction version C (3.9cm×5.7cm)

4 participants said to be not easy to understand, 6 participants said to hard to understand .It was unpopular because of too small.

We performed a one-way ANOVA(repeated-measures Design). The expected main effect of size of tactile paintings was significant, $F(2,18)= 49.78, p<.01$). Result of Multiple comparison.

(3)comparison of plaster material with reduced material(B5 size)

Average point of recognition of the plaster material was 3.4(SD 0.52),while on the other hand , average point of the reduced material was 3.3(SD 0.48). The significant difference was not detected with the ANOVA.

As shown in fig.4, there is little to distinguish small size from original size in recognition of the materials shape.

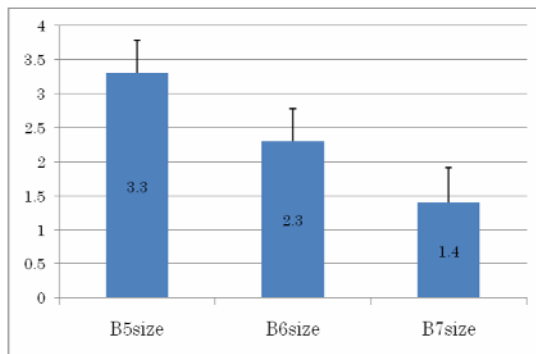


Fig. 4. Recognition of reduced materials in size of relief painting “Mona Lisa”

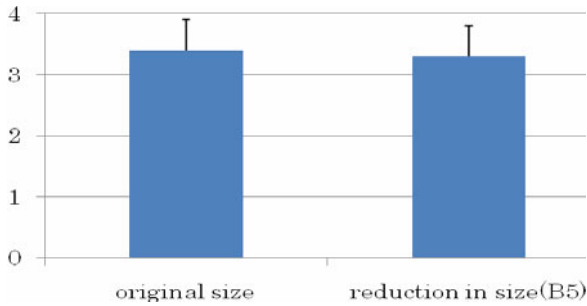


Fig. 5. Comparison of plaster material with reduced material(B5 size)

2.5 Discussion

Judging by the results of this experiment, about the relief of “Mona Lisa” used in this study, even at the reduction the size to about B5 size (version A) Blind participants were able to understand the three-dimensional image with tactile perception. If they appreciate the plaster material of relief painting “Mona Lisa” closely, they will expand their understanding of three-dimensional figure. In case of the material of B7 size(version B), recognition of the material in detail was uncertainly, but it’s possible to recognize rough configuration of the material. In contrast, in case of the material of B9 size(version C), we recognized in inappropriate materials for tactile appreciation too small.

3 Conclusion and Future Tasks

We developed reduced materials in size of relief tactile painting that blind person review what was participated original tactile painting. We confirmed the possibility of the use of the materials. We featured “Mona Lisa” in this study, the result suggest that reduced material of B7 size can use as a material for later review, but B9 size is unsuitable for tactile appreciation.

Reduction ratio of the material is affected by complexity of three dimensional shapes of that.

As measuring method of the complexity, we focus attention on degree of surface change per unit volume as shown in fig. 2. From this standpoint, we consider about method for discriminating the materials size corresponding to complexity of three-dimensional surface.

Acknowledgements

This study was partially supported by Grant-in-Aid for Scientific Research (A) (18200049) for Japan Society for the Promotion of Science(JSPS).

References

1. Aleman, A., van Lee, L., Mantione, M.H.M., Verkoijen, I.G., de Haan, E.H.F.: Visual imagery without visual experience: evidence from congenitally totally blind people. *NeuroReport* 12, 2601–2604 (2001)
2. Anderson, J.R.: Arguments concerning representations for mental imagery
3. *Psychological Review* 85, 249–277 (1978)
4. De Beni, R., Cornoldi, C.: Imagery limitations in totally congenitally blind subjects. *Journal of Experimental Psychology: Learning, Memory and Cognition* 14, 650–655 (1988)
5. Heller, M.A.: Tactile memory in sighted and blind observers: the influence of orientation and rate of presentation. *Perception* 18, 121–133 (1989)
6. Kennedy, J.M.: *Drawing and the Blind Picture to touch*. Yale university press, New Haven (1993)
7. Oouchi, S.: Utilizing a 3D scanner to produce and reproduce educational materials to learn “haptic Paintings” for persons with visual impairments. In: *Proceedings of the 30th sensory substitution symposium*, pp. 127–132 (2004) (in Japanese)
8. Secchi, L.: *L’ educazione estetica per l’ integrazione*. Carocci, Rome (2004)
9. Zimler, J., Keenan, J.M.: Imagery in the congenitally blind: How visual are visual images? *Journal of Experimental Psychology: Learning, Memory and Cognition* 9, 269–282 (1983)

Tactile Map Automated Creation System to Enhance the Mobility of Blind Persons—Its Design Concept and Evaluation through Experiment

Kazunori Minatani¹, Tetsuya Watanabe², Toshimitsu Yamaguchi², Ken Watanabe²,
Joji Akiyama², Manabi Miyagi³, and Susumu Oouchi⁴

¹ National Center for University Entrance Examination Komaba 2-19-23,
Meguro-ku, Tokyo, 153-8501 Japan

² Faculty of Engineering/Graduate School of Science and Technology,
University of Niigata Ikarashi-2 8050, Nishi-ku, Niigata 950-2181 Japan

³ Tsukuba University of Technology Kasuga 4-12-7, Tsukuba, Ibaraki, 305-0821 Japan

⁴ National Institute of Special Needs Education Nobi 5-1-1, Yokosuka,
Kanagawa, 239-0841 Japan

minatani@rd.dnc.ac.jp,
{t2.nabe,t.yamaguchi}@eng.niigata-u.ac.jp,
{f10c125j,f10c088a}@mail.cc.niigata-u.ac.jp,
mmiyagi@k.tsukuba-tech.ac.jp,
oouchi@nise.go.jp

Abstract. The authors have developed a tactile map creation system (TMACS). It is intended to assist blind persons' independent mobility. For this purpose, the system was designed to produce tactile maps, to be manipulated by the blind person and to support producing tactile maps of arbitrary locations of Japan. Through group interviews with blind persons, the authors collected information on what kind of strategies are useful for independent walk and what kind of objects can function as landmarks. TMACS is developed to make good use of these information. From the walking experiment, some assumptions which were made by the authors were confirmed. On the other hand, some unexpected or contradicted results were observed on the usefulness of landmarks and the cause of losing right routes.

Keywords: Blind person, Visually Impaired, Tactile Map, Independent Mobility, Orientation and Mobility.

1 Introduction

Tactile maps are helpful for blind person's mobility. The mobility of blind persons is considerably constrained by the fact that they can not easily understand geographical relationships, in particular, between the start and goal point. Certain degree of this constraint can be alleviated by using tactile maps. For independent mobility of blind persons, tactile maps should be generated by themselves because it is necessary for them to be able to examine their walking route anytime freely. Therefore, tactile map

creating systems must be manipulated by blind persons without any assistance of sighted persons.

Up until now, no such system which fulfills above requirements exists in Japan. Geographical Survey Institute of Japan have developed a tactile map production system which works on Microsoft Windows platform. This system was named Shokuchizu Genkou Sakusei System[1]. With that system, arbitrary locations of Japan can be printed on capsule paper as tactile maps. But that system is totally based on a GUI application, a selection of objects to be included in tactile maps are operated with the mouse. Thus, a blind person who manipulates the computer using a screen reader can not use this system.

Tactile Map Automated Production system (TMAP) is operated in the United States[2]. With this system, users can print out tactile maps of arbitrary locations in the USA with a braille (graphic) embosser. This system is designed for blind persons to be able to manipulate by themselves. Authors are influenced by the design concept of TMAP, however, this system supports only maps of the USA, so it can not produce tactile maps of locations in Japan.

Using Web applications like Google Maps[3], sighted persons can consult the map anytime freely. However such applications can neither produce tactile maps nor be manipulated by the blind person.

Therefore, the authors developed a system which satisfies above requirements and named it Tactile Map Automated Creation System (TMACS)[4].

Table 1. Comparison of Map Creation Systems

product name	Producible tactile maps?	Manipulatable by the blind person?	Compatible in Japan?
Google Maps	no	No	yes
Shokuchizu Genkou Sakusei System	yes	No	yes
TMAP	Yes	Yes	no
TMACS	yes	Yes	Yes

It should be noticed that the authors' TMACS is an attempt to assist blind persons' mobility being based on researches about the readability of tactile graphics and the training of tactile observation. Tessa Wright et. al. reviewed such basic researches published between 1965 and 2008[5]. Then it was affirmed that certain achievement had been accomplished on educational matters of tactile recognition of the (young) blind person. Meanwhile, they claimed to effect on blind persons' practical mobility by achievements of such basic researches. Maps which are generated by TMACS are designed to be highly readable in accordance with prior researches and guidebooks which were compiled with academic insights[6][7][8].

Blind persons in Japan will expand their mobility if the authors' system is fully realized. In Japan, almost all roads in urban areas are paved with asphalt and have few bumps or holes. Guide blocks (tactile ground surface indicators) which can be recognized by the feet even through the soles of shoes are laid around and inside of public buildings and stations. Therefore, it is probable that in Japan if the blind person can easily understand geographical relationships, he/she can walk comparatively safely without any assistance from sighted persons. TMACS is intended to furnish such geographical information which are necessary for blind persons' independent mobility.

Also TMACS is to be of benefit to blind persons who do not or can not walk out alone. By reading tactile maps generated by TMACS, they can understand routes which they will pass with a guide assistant. Having the knowledge on the route beforehand, they can consider the route on which are now walking. Then, they can acquire confidence on their orientation even when guided by other people. It is expected that a series of such experiences improves their ability of orientation and mobility.

Originally TMACS was developed to be utilized by Japanese blind persons. It was, however, designed to be applicable to any other regions than Japan. TMACS works on Apache 2.2.9 Web server[9] with Mapserver 4.10[10] on Linux (Fedora release 8) operating system[11]. Map data for TMACS is stored as GIS data of Mapserver. Therefore if GIS data of any country are accommodated, TMACS can serve as a tactile map production system for that country with the smallest effort.

2 Information and Objects Which Should Be or Should Not Be Included in a Tactile Route Map

Much information included in maps for sighted person must be omitted positively as the human ability of tactile recognition is considerably restricted compared with the visual recognition. Therefore information which is indicated on usual maps can not be fully included in tactile maps for blind persons. Plus, maps for sighted persons include objects which are only recognized visually (ex. signs hanged out on the roof, and colors of buildings).

Intersections and alleyways should appear on a tactile map as exactly as possible, for they are essential to plan a route to the destination[12]. When blind persons plan a course to walk referring a tactile map, generally they count alleyways which are crossing the road they are walking along and design a plan such as "go northward on the slightly winding road passing five alleyways, then turn eastward at the intersection and go straight passing two alleyways". So intersections and alleyways are essential in a tactile map.

When a blind person recognizes the circumstance of an area, some objects also play a particularly important role. While walking on the route, recognizing such objects allow blind persons to check whether or not his/her understanding of the present place is correct.

At three workshops on tactile maps, authors conducted group interviews with blind persons as to what kind of objects were useful as blind person's landmarks while walking independently. The first workshop on July 2008 was attended by 7 blind person, the second on February 2009 was 5, third on January 2010 was 8. Their answers can be summarized as follows.

1. Gas stations: Generally they locate at the corner of a block and two sides of a structure which adjacent to crossing roads are widely opened for customers' car entrance. They often smell of gasoline. So, blind persons can easily identify them via his/her auditory and olfactory senses.

2. Convenience stores: In Japan, in particular in urban areas, a man can find many shops called "convenience stores". They are comparatively small shops and are open around the clock. When their automatic door opens, unique sounds are often heard, such as an electric cash registers and background music with signage announcement. These stores usually sell cooked and ready-to-eat meals at the counter. So the stores smell of these foods, as well. These characteristic are recognized by acoustic or olfactory senses.

3. Fast food restaurants: They have similar characteristics to convenience stores.

4. Pinball saloons (Pachinko parlor): In Japan, a man can find a few pinball saloons in front of almost all urban railway stations. They are making enormous amount of noises; electric sound effects, jingles and announcements at full volume inside a hall. In some cases, their noises are stressful for a walking blind person by blocking environmental sounds, however they can be heard with no concentration, so they are functioning as auditory landmarks.

3 The Design Concept of the Tactile Map Automated Creation System

Acoustic and olfactory landmarks should be placed on the tactile map for blind persons' independent walk.

On the other hand, through interviews, it also became clear that the recognition of landmarks shows wide differences depending on occasions and from person to person. All landmarks are not recognized perfectly all the time. Landmarks which are used as cues for a blind person are different among individuals. Landmarks which are described above are not available in anytime, for instance shops do not make sounds or spread smells when they are closed. These facts should be taken into account in designing TMACS. Based on this understanding, the system has been developed so that its users can select landmarks arbitrarily.

These insights should be considered to be applicable to any tactile map production system for blind persons which supports any region. Landmarks which function as acoustic or olfactory landmarks vary according to cultures. At the same time, it is immutable that objects and places with acoustic or olfactory characteristics function as landmarks for the blind person. Thus, for the development of a tactile map production system for blind persons which supports arbitrary regions, it is important that available GIS data should not only contain many landmarks as possible but also subdivide them into categories as possible. In general, landmarks contained in GIS data for sighted persons are not categorized by acoustic or olfactory characteristics. Therefore, system developers must contrive a way to automatically select useful landmarks for blind persons out of general GIS data.

4 Experiments Conducted for the Evaluation of Tactile Map Automated Creation System

To evaluate the effectiveness of tactile maps generated by TMACS in assisting blind persons' walk, the authors conducted a walking experiment.

Procedure

Blind participants were required to walk to the destination which is less than 1 km away from the starting point. Tactile maps generated by TMACS on the capsule paper were provided to the participants. The maps pointed out the starting point, the destination and landmarks which were considered to be useful for blind person's walking. The participants read the tactile maps and planned the route from the starting point to the destination. Areas at which experiments were carried out were selected from locations where participants have never visited before. So participants could not rely on their memory but had to utilize the tactile maps.

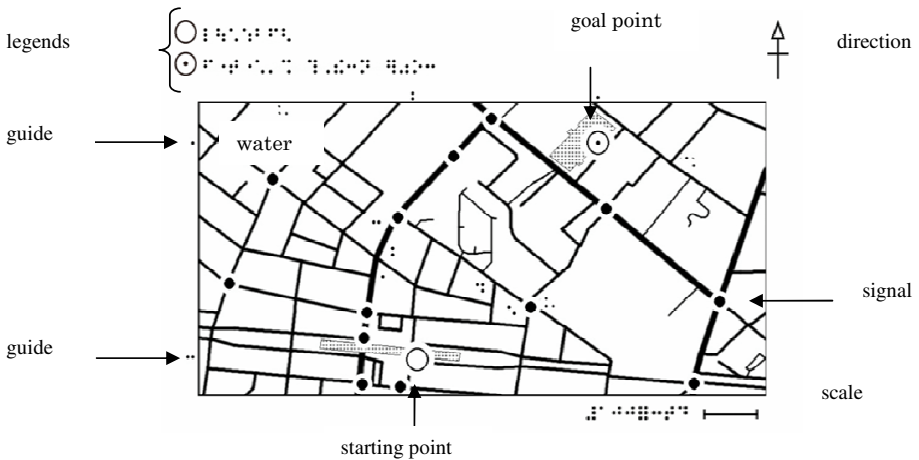


Fig. 1. The Tactile Map of Route 1

Experimental areas were selected out of Tokyo with the condition that the area has listed landmarks within its radius of 300 m around a station: more than ten signals, more than six convenience stores, more than four fast food restaurants and at least a gas station, a drug store and a pinball saloon. Authors could not find areas which fully met above conditions. Then, most desirable ones were selected with consideration on convenience to carry out an experiment. Selected areas are shown below.

Route 1: from Nishi-kasai station to Edogawa sports center (700m) (Figure 1)

Route 2: from Monzennakachou station to Tomioka-Hachiman shrine (420m)

Route 3: from Hacchoubori station to Tokyo Stock Exchange (700m)

Two participants were requested to walk along these three routes. Indeed, it was carried out with only two participants. This experiment required of a participant a certain

degree of ability of mobility. If he/she did not have such ability, participants would put themselves in danger. So, the authors selected participants very cautiously, and the number of participants were limited. The number of participants should be increased in the future.

To ensure the security of participants, three assistants observed the environment; two both in front and behind of the participant distantly and another one following the participant closely to tell critical information on security---for example signals color---

5 Results of the Experiment and the Conclusion

Results of the experiment is shown below (Table 2 and 3).

Table 2. Time Required for Reading Maps and Walking Routes (unit: minutes)

	Time for reading tactile map		Time for walking	
	participant A	participant B	participant A	participant B
route 1	11.25	12.78	17.83	21.42
route 2	7.33	7.32	16.50	11.25
route 3	6.75	7.28	17.38	18.93

Table 3. Recognition Rate of Landmarks

	participant A				participant B			
	route 1	route 2	route 3	rate	route 1	route 2	route 3	rate
intersection (pass)	3/4	6/7	10/10	90	3/4	7/7	10/10	95
intersection (turn)	1/1	1/2	2/2	80	1/1	1/1	2/2	100
Signal	3/4	0/0	4/4	88	2/4	0/0	4/4	75
convenience store	0/2	0/1	0/2	0	0/2	0/1	0/2	0
fast food restaurant	0/1	0/0	0/1	0	0/1	0/0	0/1	0
gas station	0/0	1/1	1/1	100	0/0	1/1	1/1	100

From the experimental results, the assumptions below were confirmed.

1. Tactile maps generated by the TMACS are useful in assisting blind persons' walk, planning a route toward the destination, and knowing the circumstances along the route.

2. The strategy of counting intersections and alleyways functions as a way to check his/her present location.

3. Gas stations can function as a landmark for blind persons.

On the other hand, results below were unexpected or contradicted with the prior assumption.

1. Convenience stores and fast food restaurants did not function as landmarks for blind persons.

2. The parking area and the entrance of public garden were misidentified as alleyways.

From these results, conclusions can be drawn below.

1. TMACS is useful in assisting blind persons' walk.

2. Some objects or places which have acoustic or olfactory characteristics do not work as landmarks in contradiction to some blind persons' comments (e.g. convenience stores and fast food restaurants).

3. Blind persons often confuse open space such as parking areas and the entrances of public gardens with alleyways. It is expected that such spaces should be signified on the tactile map.

Acknowledgments. This work was partially supported by Grant-in-Aid for Scientific Research (B) (20300200) of Japan Society for the Promotion of Science(JSPS).

References

1. <http://zgate.gsi.go.jp/shokuchizu/>
2. Miele, J., Marston, J.: Tactile Map Automated Production (TMAP): Project Update and Research Summary. In: Proceedings CSUN 2005 (2005)
3. Watanabe, T., Yamaguchi, T., Minatani, K., Oouchi, S., Miyagi, M., Iwashita, Y.: Development of a Tactile Map Production System Accessible to Blind Persons. In: IEICE technical report, vol. 108(488), pp. 13–18. IEICE, Tokyo (2009)
4. <http://maps.google.com>
5. Wright, T., Haris, B., Sticken, E.: A Best-Evidence Synthesis of Research on Orientation and Mobility Involving Tactile Maps and Models. *Journal of Visual Impairment & Blindness*, 95–106 (February 2010)
6. Eriksson, Y., Jansson, G., Strucel, M.: Tactile maps - Guidelines for the production of maps for the visually impaired. The Swedish Library of Talking Books and Braille, Stockholm (2003)
7. Eriksson, Y., Strucel, M.: A Guide to the Production of Tactile Graphics on Swellpaper. The Swedish Library of Talking Books and Braille, Stockholm (1995)
8. Edman, P.: Tactile Graphics. AFB Press, Huntington (1992)
9. <http://www.apache.org>
10. <http://mapserver.org/>
11. <http://fedoraproject.org/>
12. Takamiya, S., Hamada, S.: Information Used by Visually Impaired People While Walking. *Transportation Research Record* 1636, 104–109 (1998)

Development of Tactile Graphics Production Software for Three-Dimensional Projections

Mamoru Fujiyoshi¹, Takeshi Kaneko², Akio Fujiyoshi³,
Susumu Oouchi², Kenji Yamazawa⁴, Yuji Ikegami⁴,
Yasunari Watanabe⁴, and Yoshinori Teshima⁵

¹ National Center for University Entrance Examinations
2-19-23 Komaba, Meguro, Tokyo, 153-8501 Japan
fujiyosi@rd.dnc.ac.jp

² National Institute of Special Needs Education
5-1-1 Nobi, Yokosuka, Kanagawa 239-8585 Japan
kaneko@nise.go.jp, oouchi@nise.go.jp

³ Department of Computer and Information Sciences, Ibaraki University
4-12-1 Nakanarusawa, Hitachi, Ibaraki, 316-8511 Japan
fujiyosi@mx.ibaraki.ac.jp

⁴ Rapid Engineering Team, RIKEN Advanced Science Institute (ASI)
2-1 Hirosawa, Wako, Saitama 351-0198 Japan
kyama@riken.jp, yikegami@riken.jp, wistajp@yahoo.co.jp

⁵ National Institute of Advanced Industrial Science and Technology
1-2 Namiki, Tsukuba, Ibaraki 305-8564 Japan
yoshinori.teshima@aist.go.jp

Abstract. In this paper, we introduce tactile graphics production software for three-dimensional projections. A blind person can use this software without assistance from the sighted and produce tactile graphics of three-dimensional projections. With this software, we want to study the limitation of tactile recognition of projections and improve the guidelines of teaching projections.

Keywords: tactile graphics, the blind, three-dimensional projections, universal design.

1 Introduction

Tactile graphics production software for three-dimensional projections is developed for the blind. A blind person can use this software without assistance from the sighted and produce tactile graphics of three-dimensional projections. The software supports various types of projections such as orthographic projection, oblique projection, and perspective projection. Once required parameters are inputted to the software, tactile graphics of three-dimensional projections will be printed by a braille plotter printer.

Although the concepts of topological space and Euclidean space are learnable through tactile sense (haptic perception) and auditory sense, it is considered

that it is almost impossible for the congenitally blind and the early blind to acquire the concept of projections because of a lack of experience in seeing light and shadow [2,3,4]. However, there might be possibility that the concept of projections becomes learnable for them when a person starts abstract thinking around age 11 or 12 years.

This study presents tactile graphics production software for three-dimensional projections that is usable by a blind person. A blind person can obtain tactile graphics of various types of projections from any preferable viewpoint. The software enables the blind to study the concept of projections by themselves. Obviously, the tactile recognition of projections is much harder than the visual recognition of projections. We want to study the limitation of tactile recognition of projections and improve the guidelines of teaching projections.

2 Development of Bplot 3D

Tactile graphics production software for three-dimensional projections Bplot 3D is developed.

2.1 Outline of Bplot 3D

Since we want to make Bplot 3D to be usable by a blind person without assistance from the sighted, it is developed as a preprocessor of the universal design tactile graphics production system Bplot2 [1].

A user gives Bplot 3D a series of commands to produces tactile graphics of a three-dimensional projection by specifying a type of projection, a viewpoint, and a projection plane. Commands are written in a text file using our newly designed control language for Bplot 3D. Because a command file for Bplot 3D is a text file, it is editable with any text editors by any person who has learned the control language. Therefore, Bplot 3D can be used by both the sighted and the blind.

The software supports various types of projections. In general, projections are classified into two categories, namely, ‘parallel projection’ and ‘perspective projection’. Moreover, ‘parallel projection’ are divided into ‘orthographic projection’ and ‘oblique projection’. In parallel projection, the projection rays from the object to the projection plane are parallel to each other. Orthographic projection is a type of parallel projection where the view direction is orthogonal to the projection plane. On the other hand, oblique projection is a type of parallel projection where the view direction is not orthogonal to the projection plane. In perspective projection, the projection rays are not parallel to each other but they are emerging from the viewpoint.

2.2 Commands for Bplot 3D

To distinguish three-dimensional commands from ordinal commands of Bplot2, every three-dimensional command starts from ‘3’.

(1) Projection Type: A type of projections needs to be specified prior to the other commands. For orthographic projection, the command is ‘3ortho-graphic’. For oblique projection, the command is ‘3oblique depth theta’. The parameters ‘depth’ and ‘theta’ specify the fraction of depth versus width, and the angle from the horizontal direction to the depth direction. For perspective projection, the command is ‘3perspective’.

(2) Viewpoint: A viewpoint is specified with the command ‘3viewpoint x0 y0 z0’. Instead of the coordinates, a viewpoint can be specified as a directional vector by the command ‘3direction theta phi’ for parallel projection.

(3) Wireframe: The wireframe of a figure is specified with the following commands. To define a plot, we use the command ‘3plot x0 y0 z0’. The command ‘3plot’ has three parameters for the coordinates of point ‘x0 y0 z0’. To define a line segment, we use the command ‘3line x0 y0 z0 x1 y1 z1 DotPitch’. The command ‘3line’ has seven parameters for the coordinates of start point ‘x0 y0 z0’, the coordinates of end point ‘x1 y1 z1’, and the dot pitch. To define connected line segments, we use the command ‘3bline’. The command ‘3bline’ specifies the dot pitch, the coordinates of start point, middle points, and end point, as follows:

```
3bline DotPitch
@x1 y1 z1
@x2 y2 y2
@x3 y3 z3
@...
@...
@...
@xn yn zn
@9999 0 0
```

On the last line, ‘9999 0 0’ means the end of command. We can also define a curve line with the command ‘3spline’.

(4) Macros: We can define macros for three-dimensional commands by using the command ‘3define identifier replacement_string_list’. The command ‘3define’ replaces ‘identifier’ into ‘replacement_string_list’ prior to the other three-dimensional commands. For the readability of three-dimensional commands, we can write parameters by identifiers instead of numerics.

2.3 An Example of Produced Three-Dimensional Projections

The following is an example of a command file to produce a oblique projection in Fig. 1. The command file ‘File 3D’ describe a wire frame of cube that has the center as the origin and sides of the length 6. The type of projection and the viewpoint are also described in the file. When ‘File 3D’ is inputted to Bplot 3D, the command file for Bplot2 ‘File 2D’ is obtained. From ‘File 2D’, Bplot2 can produce tactile graphics with a braille plotter printer.

On line 5 and 6 in ‘File 3D’, a type of projections and a viewpoint are specified. These lines are commented out in ‘File 2D’. On lines 8-10 and 50-51 in ‘File 3D’,

File 3D

```

1: /* Wire Frame of Cube
2:   by Oblique Projection
3: */
4:
5: 3oblique 0.67 45
6: 3viewpoint 0 -10 0
7:
8: 3define dotpitch1 2.2
9: 3define dotpitch0 4.0
10: 3define end_of_data 9999 0 0
11:
12: define small 0
13: define middle 1
14: define large 2
15:
16: /// visible lines
17:
18: dot middle
19: 3bline dotpitch1
20:   -3.0 -3.0 3.0
21:   3.0 -3.0 3.0
22:   3.0 3.0 3.0
23:   -3.0 3.0 3.0
24:   -3.0 -3.0 3.0
25:   end_of_data
26:
27: 3bline dotpitch1
28:   -3.0 -3.0 3.0
29:   -3.0 -3.0 -3.0
30:   3.0 -3.0 -3.0
31:   3.0 -3.0 3.0
32:   end_of_data
33:
34: 3bline dotpitch1
35:   3.0 -3.0 -3.0
36:   3.0 3.0 -3.0
37:   3.0 3.0 3.0
38:   end_of_data
39:
40: /// hidden lines
41:
42: dot small
43:
44: 3bline dotpitch0
45:   -3.0 -3.0 -3.0
46:   -3.0 3.0 -3.0
47:   3.0 3.0 -3.0
48:   end_of_data
49:
50: 3define point0 -3.0 3.0 -3.0
51: 3define point1 -3.0 3.0 3.0
52: 3line point0 point1 dotpitch0
53:
54: ///
55:
56: ff // formfeed

```

File 2D

```

1: /* Wire Frame of Cube
2:   by Oblique Projection
3: */
4:
5: //3D oblique projection (0.67, 0.25Pi)
6: //3D viewpoint (0.00, -10.00, 0.00)
7:
8: //3D defined dotpitch1 2.2 in rec. 1
9: //3D defined dotpitch0 4.0 in rec. 2
10: //3D defined end_of_data 9999 0 0 in rec. 3
11:
12: define small 0
13: define middle 1
14: define large 2
15:
16: /// visible lines
17:
18: dot middle
19: bline 2.200 //3D 5 points
20:   -4.421 1.579 //3D point 1
21:   1.579 1.579 //3D point 2
22:   4.421 4.421 //3D point 3
23:   -1.579 4.421 //3D point 4
24:   -4.421 1.579 //3D point 5
25:   9999 0 //3D end_of_data
26:
27: bline 2.200 //3D 4 points
28:   -4.421 1.579 //3D point 1
29:   -4.421 -4.421 //3D point 2
30:   1.579 -4.421 //3D point 3
31:   1.579 1.579 //3D point 4
32:   9999 0 //3D end_of_data
33:
34: bline 2.200 //3D 3 points
35:   1.579 -4.421 //3D point 1
36:   4.421 -1.579 //3D point 2
37:   4.421 4.421 //3D point 3
38:   9999 0 //3D end_of_data
39:
40: /// hidden lines
41:
42: dot small
43:
44: bline 4.000 //3D 3 points
45:   -4.421 -4.421 //3D point 1
46:   -1.579 -1.579 //3D point 2
47:   4.421 -1.579 //3D point 3
48:   9999 0 //3D end_of_data
49:
50: //3D defined point0 -3.0 3.0 -3.0 in rec. 4
51: //3D defined point1 -3.0 3.0 3.0 in rec. 5
52: line -1.579 -1.579 -1.579 4.421 4.000 //3D
53:
54: ///
55:
56: ff // formfeed

```

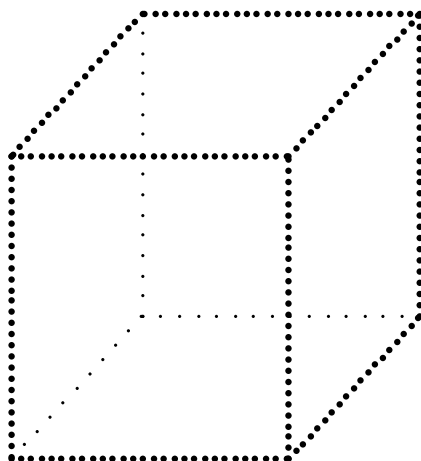


Fig. 1. Oblique projection of a cube

macros are defined. Parameters of other commands are specified by identifiers instead of numerics. These lines are commented out in 'File 2D'. On lines 12-14 in 'File 3D', there is the command 'define'. Since this command is not a three-dimensional command, Bplot 3D does not touch it. On line 18 and 42 in 'File 3D' and 'File 2D', the size of dots of dotted lines is specified with the command 'dot'. On lines 16-38 in 'File 3D' and 'File 2D', the wire frame of a cube on the visible side is defined. On lines 40-52 in 'File 3D' and 'File 2D', the wire frame on the hidden side is defined. On line 56 in 'File 3D' and 'File 2D', the command 'ff' means the end of page.

3 Experimental Evaluation

In order to evaluate the usefulness of Bplot 3D, an experiment was conducted to study which type of projection is suitable for blind students to recognize shapes of three-dimensional objects.

As experimental materials, we prepared tactile graphics of projections of four kinds of geometric solids (cube, cuboid, cylinder, and tetrahedron) by the following drawing methods: (1) development view, (2) front view of orthographic projection, (3) oblique projection without hidden lines, (4) oblique projection with hidden lines, and (5) perspective projection with hidden lines. An example of five tactile graphics of a cube is shown in Fig. 2.

Each figure was embossed on a sheet of braille-printer paper. The visible lines of wire frames were drawn by thick dotted lines (dot size 1.5mm in diameter and dot pitch 2.2mm) and hidden lines were delineated by thin lines (dot size 0.8mm in diameter and dot pitch 4.0mm).

Subjects were eight totally blind students in a high school for the blind. Their age ranged from 16 to 17 years old. They were classified into two subject groups:

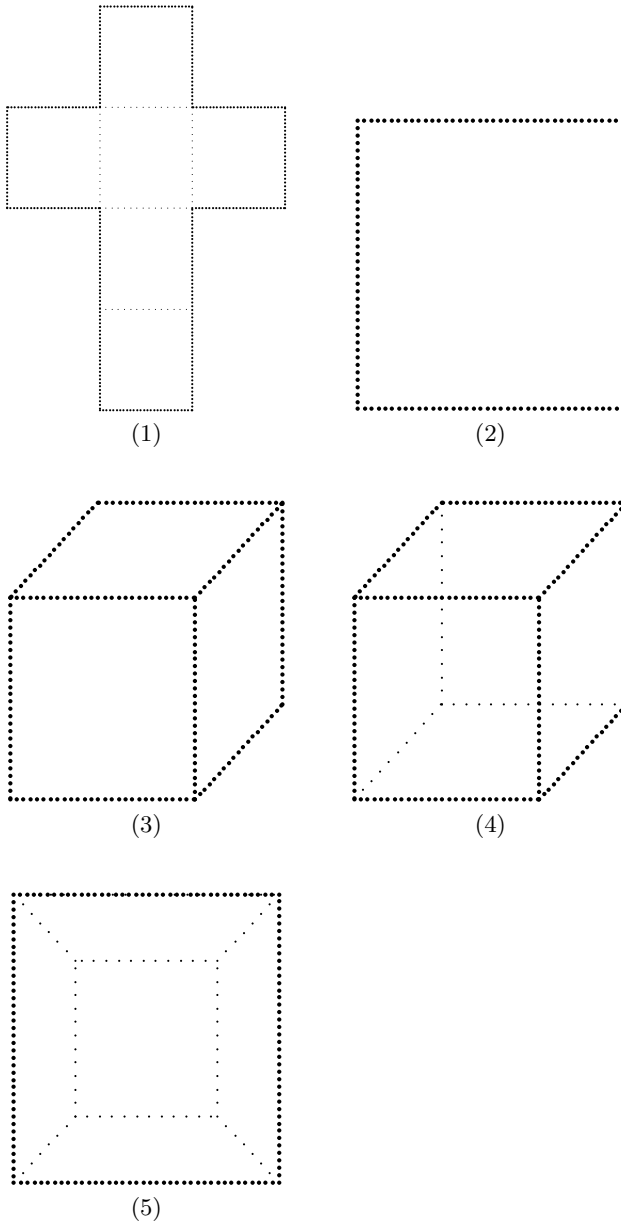


Fig. 2. Tactile graphics of a cube by the five drawing methods

the early blind and the late blind. In the early blind group, they suffered visual impairments before 5 years old. In the late blind group, they suffered visual impairments from 6 to 11 years old.

The subjects were asked to succeedingly observe tactile graphics of one of the four geometric solids by one of the five drawing methods. Simultaneously, they

were required to rank the five drawing methods. The task was repeated for the four geometric solids, respectively.

As a result of the experiment, we found that blind students prefer development view as the best drawing method. Oblique projections with and without hidden lines are at the second and third, and front view of orthographic projection and perspective projection are at the fourth and fifth. From the result of analysis of variance and Scheffe's comparison, the ranks of the five drawing methods were significant for all four geometric solids.

We also found that subjects in the late blind group recognized the concept of projection better than subjects in the early blind group through discussion with them. However, we could not find significant differences between two subject groups because the number of subjects was too small.

4 Conclusion

Tactile graphics production software for three-dimensional projections Bplot 3D was developed for the blind. Bplot 3D was implemented as a preprocessor of the universal design tactile graphics production system Bplot2 [1]. Both the blind and the sighted can produce tactile graphics of projections easily.

As a result of the experiment, we noticed that it is not easy for the blind to recognize projections. However, the study of projection is necessary in natural sciences for higher education [2,3,4,5].

To know the limitation of tactile recognition of projections and to improve the guidelines of teaching projections, we need further studies.

Acknowledgement

This work was partially supported by Grant-in-Aid for Scientific Research (A) (18200049) of Japan Society for the Promotion of Science(JSPS).

References

1. Fujiyoshi, M., Fujiyoshi, A., Ohtake, N., Yamaguchi, K., Teshima, Y.: The Development of a Universal Design Tactile Graphics Production System Bplot2. In: Miesenberger, K., Klaus, J., Zagler, W.L., Karshmer, A.I. (eds.) ICCHP 2008. LNCS, vol. 5105, pp. 938–945. Springer, Heidelberg (2008)
2. Heller, M.A., Brackett, D.D., Scroggs, E., Steffen, H., Heatherly, K., Salik, S.: Tangible pictures: Viewpoint effects and linear perspective in visually impaired people. *Perception* 31, 747–769 (2002)
3. Kennedy, J.M., Gabias, P., Nicholls, A.: Tactile Pictures. In: Heller, M.A., Schiff, W. (eds.) *The Psychology of Touch*, pp. 263–299. Lawrence Erlbaum, Hillsdale (1991)
4. Kennedy, J.M.: Drawings from Gaia, a blind girl. *Perception* 32, 321–340 (2003)
5. Kennedy, J.M., Juricevic, I.: Haptics and projection: Drawings by Tracy, a blind adult. *Perception* 32, 1059–1071 (2003)

Comprehending and Making Drawings of 3D Objects by Visually Impaired People: Research on Drawings of Geometric Shapes by Various Methods of Projection*

Takeshi Kaneko¹, Mamoru Fujiyoshi², Susumu Oouchi¹, Yoshinori Teshima³, Yuji Ikegami^{3,4}, Yasunari Watanabe^{3,4}, and Kenji Yamazawa⁴

¹ National Institute of Special Needs Education, 5-1-1 Nobi, Yokosuka, Kanagawa 239-8585 Japan

² Organization of the Study for College Admissions, National Center for University Entrance Examinations, 2-19-23 Komaba, Meguro, Tokyo 153-8501 Japan

³ Advanced Manufacturing Research Institute, National Institute of Advanced Industrial Science and Technology, 1-2 Namiki, Tsukuba, Ibaraki 305-8564 Japan

⁴ Rapid Engineering Team, RIKEN Advanced Science Institute(ASI), 2-1 Hirosawa, Wako, Saitama 351-0198 Japan

{Kaneko,oouchi}@nise.go.jp, fujiyosi@rd.dnc.ac.jp,
yoshinori.teshima@aist.go.jp, {yikegami,kyama}@riken.jp,
wistajp@yahoo.co.jp

Abstract. In this study, we investigated the possibility of the early and late blind comprehending and producing tactile drawings by oblique, axonometric or perspective projection, which has been considered difficult to understand for them, especially for the early blind. For this purpose, an experiment was carried out analyzing the following issues: 1)how early and late blind high school students draw four geometric 3D shapes; 2)how they rank tactile drawings of such shapes produced via various methods; development, orthogonal projection, in addition to above three projection methods; 3)their explanations of how these drawings by use of such methods are produced. The results demonstrate that both groups understand drawings via latter two methods well and in addition to the late blind, even the early blind understand drawings via the former three projection methods at least partially, which would lead to better understanding.

Keywords: Visually Impaired People, Drawing of 3D Object, Tactile Recognition, Educational Material.

1 Introduction

There are methods of drawing that represent 3D objects on a plane, such as oblique projection, axonometric projection and perspective projection, which are considered easy to understand, visually. Drawings produced by perspective projection, or 'perspective,' closely resemble real 3D objects, so they are easy to understand for the

* This research was partially supported by Grant-in-Aid for Scientific Research (A) (18200049) of Japan Society for the Promotion of Science (JSPS).

sighted. Drawings produced by oblique projection and axonometric projection also resemble real objects and are effective methods of conveying information on true shape and shape components. As to examples of drawings made by these methods, refer to Fig. 2. However, blind people seem to have difficulty in interpreting drawings and depictions of 3D objects produced via these methods, as 2D representation of 3D objects is deeply dependent on visual perception. This idea is based upon the results of previous studies of how blind individuals comprehend 2D representations of 3D objects, and the pictures such individuals produce of 3D objects [1], [2], [3], [4], [5], [6], [7], [8]. These studies include the results that the blind are incapable of understanding drawings created via the above methods, and that they draw 2D pictures of 3D objects via different methods, employing orthogonal projection or development. These studies, however, also include some results that call this conclusion into question. In particular, it appears that individuals that have become blind later in life are more likely to exhibit these results than the congenitally blind or individuals that have lost their sight early in life. Besides, there are also some cases of the congenitally or early blind that oppose this conclusion [5], [6], [7], [8]. Yet, in such results, various methods of projection, mentioned above, are not always distinguished explicitly. In addition the presence of hidden lines in drawings produced via such methods is not always considered.

Blind people, especially those that are congenitally blind, or that lost their sight early in life, might not be able to understand drawings produced via the method of oblique, axonometric, or perspective projection in strictly the same manner as a sighted individual. It is also possible that they would find it difficult producing drawings by such methods. However, it is possible that they might still understand these methods conceptually, or mathematically, if they learn them by way of educational processes that focus upon tactile sensation, employing models of 3D shapes, planes of projection, lines of projection, eye point, etc. as their components. In actuality, one of the authors in the present study, Kaneko, produced such educational materials and began tentative use of them to educate the blind. Refer to Fig.1 as the examples of such educational materials. In addition, drawings of 3D objects have recently been made easier to produce with the use of some computer programs. Such programs require only the setting of types of projections and certain parameters: coordinate values of 3D shapes, the eye point, the direction of projection lines, slant values of shapes, etc. Based upon this technology, another author, Fujiyoshi, developed a 3D drawing system for the blind, referred to as 'Bplot 3D,' which enables the blind to independently produce drawings on a computer and to print tactual drawings with the use of Braille printer [9]. By using such systems, the blind are made capable of producing 3D drawings quite easily.

The present study aims to assess the possibility of the blind recognizing and producing drawings by oblique, axonometric or perspective projection, which has heretofore been considered quite difficult. For this purpose, an experiment was carried out analyzing the following issues: 1)how the congenitally blind or individuals who lost their sight early in life and those who lost their sight late in life draw four geometric 3D shapes (cube, cuboid, cylinder and trigonal pyramid); 2)how they rank tactile drawings of such shapes produced via various methods; development, orthogonal, oblique, axonometric and perspective projection, with or without hidden lines; 3)their explanations of how these drawings by use of such methods are produced.

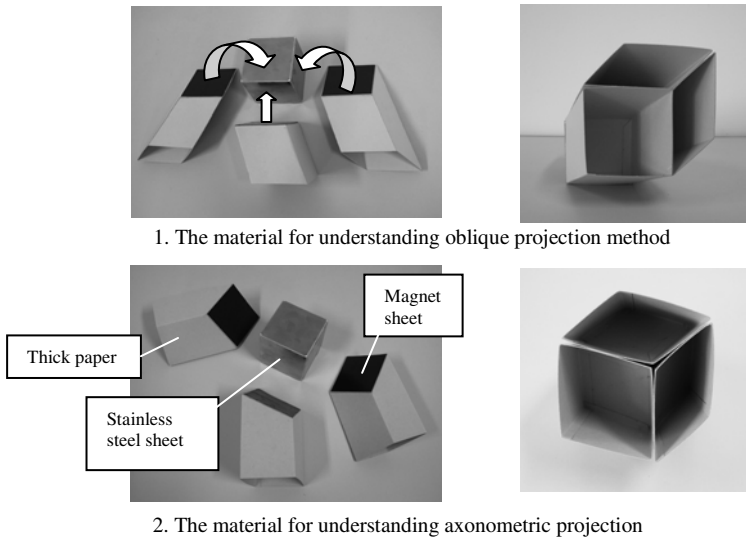


Fig. 1. Examples of educational materials for understanding projection methods tactually. In both (1) and (2), the components of the material shown in the left are assembled as shown in the right. The cube under stainless steel sheets is woodwork and 6.0×6.0×6.0 cm. Parts attached to the cube are made by development of thick paper, with magnet sheets, so they could easily be made by everyone, including teachers for visually impaired students.

2 Method

Four high school students that became blind early in their lives (early blind) and four that became so later in life (late blind) took part in the experiment. The subjects were selected from a school for the blind. Both groups were aged between 16 and 17 years (the mean age being 16 years and 3 months). The early blind students suffered from visual impairment between the ages of 0 and 5 years and began reading Braille between the ages of 5 and 7 years. The late blind students suffered from visual impairment between the ages of 7 and 11 years and began reading Braille between the ages of 10 and 14 years. Both groups of subjects had a visual acuity ranging from no light sense to sense of hand movements in front of the eye.

The following three procedures, in order, were carried out for each subject.

1) Subjects were required to draw the following 4 geometric shapes, exploring them tactually, by use of a Japanese, raised-line drawing kit: cube (6.0 cm), cuboid (4.5×4.5×9.0 cm), cylinder (6.0 cm diameter×6.0 cm height), trigonal pyramid (regular tetrahedron) (7.3×6.7×6.0 cm height). The material comprising the shapes was thick paper. Subjects were required to explain each drawing after completion.

2) Subjects were required to rank the tactile drawings produced by the different methods, including their own drawing (a) made in procedure 1), in terms of representing each of the 4 above geometric shapes well. In addition to drawing (a), we prepared tactile drawings via the following 8 methods: development (b), orthogonal

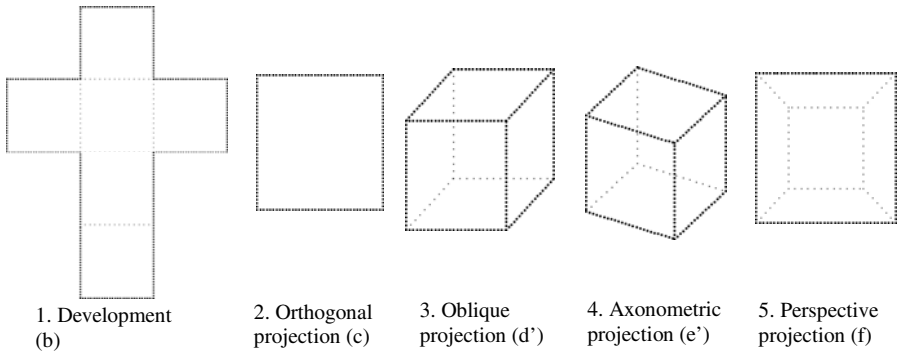


Fig. 2. Examples of tactile drawings used for the cube. The drawings (3), (4) and (5) have hidden lines. Among these projection methods, orthogonal projection, oblique projection and axonometric projection have projection lines parallel to each other, so they are called parallel projection. Besides, in orthogonal projection, a plane of the object is parallel to and projection lines are orthogonal to the plane of projection. In oblique projection, a plane of the object is parallel to and projection lines are oblique to the plane of projection. In axonometric projection, planes of the object are oblique to and projection lines are orthogonal to the plane of projection. In perspective projection, projection lines aren't parallel to each other, but converge at one point through the plane of projection. This point is eye point.

projection without hidden lines (c), orthogonal projection with hidden lines (c'), oblique projection without hidden lines (d), oblique projection with hidden lines (d'), axonometric projection without hidden lines (e), axonometric projection with hidden lines (e') and perspective projection (with hidden lines) (f). Examples of these tactile drawings for the cube are shown in Fig. 2. Drawing (c) had front view only with the exception of the trigonal pyramid. As to the trigonal pyramid, drawing (c) had bottom view only. Drawing (c') was produced for the trigonal pyramid and had bottom view only with hidden lines representing its side faces. Drawing (f) was produced with the shapes' faces parallel to the plane of projection. The shapes' parallel faces were the front for the cube and cuboid, the top for the cylinder. Of these 9 types of drawings, (a) through (f), the subjects were randomly presented with 8 for the cube and cuboid (all but c'), 6 for the cylinder (a, b, c, e, e', and f) and 6 for the trigonal pyramid (a, b, c, c', d and d'). Subjects were required to rank these drawings and then explain their reasoning. The prepared tactile drawings were produced using 'Blpot 3D' and printed using a Braille printer (ESA721). Solid lines were made of embossed dots, 1.5 mm in diameter, at 2 mm intervals. Dotted lines depicting development folding lines or hidden lines were made using dots 0.8 mm in diameter, at 4 mm intervals.

3) With regard to the cube, subjects were presented with 7 kinds of drawings (b, c, d, d', e, e', and f), in order, and asked to attempt to explain how each drawing was produced.

3 Results and Discussion

Drawings made by subjects in procedure 1) were categorized as one of type (b) through (f) used in procedure 2) or, for drawings which did not fall within one of these

categories, into a new classification, such as (c+), (c''), based upon their characteristics. Types of drawings produced in procedure 1) and ranked highest in procedure 2) by each subject are shown in Table 1. The number of types of drawings in procedure 1) and ranked highest in procedure 2) by early and late blind subjects for each shape are shown in Table 2. The examples of drawings made by subjects are shown in Fig. 3.

Table 1. Types of drawings produced in procedure 1) and ranked highest in procedure 2) for each 3D shape by each subject. As to meanings of symbols in the table, refer to the text.

Subject		Early Blind								Late Blind							
		A		B		C		D		E		F		G		H	
Shape	Procedure	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2
Cube		b	b	c+	b	b	b	c	b	e	d	e'	e'	d	e	cα	b
Cuboid		b	b	c+	b	b	b	c	b	e	d	d'	e'	d	d	cα	b
Cylinder		b	b	c+	a	b	b	c+	b	e'	e'	b	b	e	e	cα	b
Trigonal Pyramid		b	b	b	b	b	b	c	b	c'	c'	b	c'	c''	b	cβ	b

Table 2. The number of types of drawings produced in procedure 1) and ranked highest in procedure 2) by early and late blind subjects for each shape. As to details of drawing methods in the table, refer to the text.

a. Procedure 1)					b. Procedure 2) (ranked highest)												
Shape	Cube		Cuboid		Cylinder		Trigonal Pyramid		Shape	Cube		Cuboid		Cylinder		Trigonal Pyramid	
Early (E) or Late (L) blind	E	L	E	L	E	L	E	L	Early (E) or Late (L) blind	E	L	E	L	E	L	E	L
Drawing method									Drawing method								
b: development	2		2		2	1	3	1	b: development	4	1	4	1	3	2	4	2
c: orthogonal projection	1		1				1		c: orthogonal projection								
c+: (c) with two views	1		1		2				c+: (c) with two views					1			
c': (c) with hidden lines							1		c': (c) with hidden lines								2
c'': (c) with solid lines							1		d: oblique projection		1		2				
cα: two connected faces		1		1		1			d': (d) with hidden lines								
cβ: one beneath two triangles							1		e: axonometric projection		1				1		
d: oblique projection		1		1					e': (e) with hidden lines		1		1		1		
d': (d) with hidden lines				1					f: perspective projection								
e: axonometric projection		1		1		1											
e': (e) with hidden lines		1				1											
f: perspective projection																	

From the results of procedure 1), early blind subjects depicted shapes by development (b) or orthogonal projection with one view only (c) or two views (c+). For procedure 2), with only one exception, one subject's depiction of the cylinder, all subjects ranked development (b) drawings highest. The outlier subject ranked his own drawing (c+) highest.

In contrast, two late blind subjects produced drawings of the cube and cuboid, using oblique projection without hidden lines (d) or using axonometric projection without hidden lines (e). One subject depicted the cube by axonometric projection with hidden lines (e') and the cuboid by oblique projection with hidden lines (d'). Two of these three subjects depicted the cylinder by axonometric projection without hidden lines (e) and with hidden lines (e') respectively. The final remaining subject produced drawings of 3 shapes with the top face connected to the front face (cα). In procedure 2), for these 3 shapes, the former three subjects ranked drawings produced by oblique or axonometric projection (d, e and e') highest, with one exception, while the final

remaining subject ranked those drawings produced by development as the highest. With regard to the trigonal pyramid, in procedure 1), one subject depicted it using orthogonal projection from the bottom with hidden lines of its side faces (c') and another subject by the same method from the top with side face's solid lines (c"). One of the remaining two subjects depicted the shape via development and the other subject, who drew the above ($c\alpha$) picture, produced drawings with one triangle beneath two joined triangles (attached at the bottom line) ($c\beta$). In procedure 2), for the trigonal pyramid, two subjects ranked drawings produced by orthogonal projection with hidden lines (c') highest, and the remaining two ranked highest those drawings produced by development (b).

In addition, neither early nor late blind subjects depicted drawings via perspective projection method (f) for each shape, nor ranked them highest.

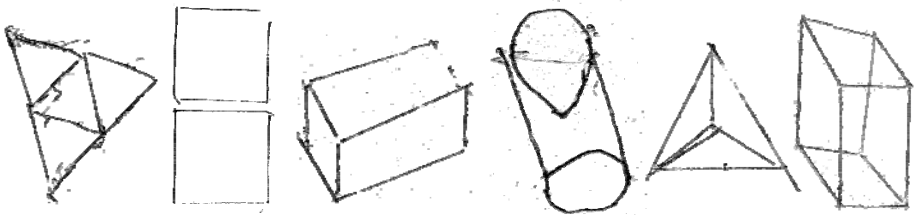


Fig. 3. The examples of drawings made by subjects. From the left, drawing produced via (b) by subject A, (c+) by subject B, (e), (e'), (c') by subject E, and (d') by subject F. These drawings were produced using a ruler.

The whole results of the rankings in procedure 2) are shown in Table 3. From the rankings of the cube and cuboid drawings in these results, two early blind subjects ranked those drawings produced by perspective projection (with hidden lines) (f) higher than ones produced by oblique projection (d, d') and axonometric projection (e, e'). In contrast, three late blind subjects ranked drawings produced by the former method lower than the latter two.

With regard to the cylinder, all of the early blind subjects ranked drawings produced by perspective projection (f) lower than drawings produced by axonometric projection (e, e'). Furthermore, drawings produced by axonometric projection with hidden lines (e') were ranked higher than those produced by axonometric projection without hidden lines (e). Drawings of the cylinder produced by axonometric projection (e, e'), being different from those of cube and cuboid, had a portion of the drawing that represented the front face of the object, with one ellipse at the top and another at the bottom (e') or one ellipse at the top and a half of ellipse at the bottom (e), while drawings produced by perspective projection were comprised of two concentric circles (f). So, the former drawings seem to be considered better drawings representing the cylinder than the latter ones by the early blind. Furthermore, it appears that (e') is better than (e) because it contains two ellipses.

The results of procedure 3) regarding the cube, were such that all subjects, both the early and late blind, were able to explain drawings exactly via development (b) and orthogonal projection with one view only (c). As to drawings produced by oblique

projection (d, d'), axonometric projection (e, e') and perspective projection (with hidden lines) (f), the late blind were almost able to explain the images exactly by use of visual expression, although they were unable to clearly recognize the difference between drawings produced by oblique projection (d, d') and by axonometric projection (e, e'). For example, one late blind subject explained on the drawing via (d) that it was seen obliquely from the right, and on the one via (d') that it had unseen parts. Other late blind subject explained on the drawing via (f) that the depth was expressed by use of dotted lines and the far face was expressed smaller than the near face. However, the early blind could not explain exactly drawings produced by these three methods, with one exception regarding perspective projection (f), mentioned below. Three late blind subjects were able to understand hidden lines in the drawings, while only one of the early blind subjects was able to understand them.

Table 3. Rankings for each 3D shape by each subject

Subject		Early Blind																															
		A								B								C								D							
Shape	Rankings	1	2	3	4	5	6	7	8	1	2	3	4	5	6	7	8	1	2	3	4	5	6	7	8	1	2	3	4	5	6	7	8
Cube		b	a	c	d	f	d'	e'	e	b	a	f	d'	d	e'	e	c	b	d	e	d'	e'	c	a	f	b	f	d'	e'	d	e	c	a
Cuboid		b	a	c	f	d	d'	e'	e	b	a	f	d	c	e'	d'	e	b	d'	e'	f	d	e	a	c	b	f	e'	d'	d	e	a	c
Cylinder		b	a	e'	e	f	c			a	b	e'	e	f	c			b	e'	e	a	f	c			b	a	e'	e	f	c		
Trigonal Pyramid		b	a	c'	d	d'	c			b	d'	a	d	e'	*			b	d'	c'	d	a	c			b	c'	d'	d	c	a		

* Drawing 'c' wasn't included here.

Subject		Late Blind																															
		E								F								G								H							
Shape	Rankings	1	2	3	4	5	6	7	8	1	2	3	4	5	6	7	8	1	2	3	4	5	6	7	8	1	2	3	4	5	6	7	8
Cube		d	e	d'	e'	b	f	c	a	e'	d'	b	a	d	e	f	c	e	d	b	c	a	d'	e'	f	b	a	d'	f	e	d	e'	c
Cuboid		d	e	d'	e'	b	f	c	a	e'	d'	b	d	e	a	f	c	d	e	a	b	c	e'	d'	f	b	d'	e'	a	d	f	e	c
Cylinder		e'	e	b	f	c	a			b	a	e'	e	f	c			e	a	b	e'	f	c			b	e'	a	e	c	f		
Trigonal Pyramid		c'	b	d'	d	c	a			c'	b	a	e'	e	c			b	c'	a	c	d	d'			b	c'	a	d'	d	c		

Even for the early blind, with regard to the drawings produced by perspective projection (with hidden lines) (f), one subject reported that the drawing had one face on the floor (representing the bottom face of the cube) in the center and had side faces which rose around the bottom face. With regard to the hidden lines, mentioned above, one early blind subject indicated that they represented 'unseen parts.' Referring to a drawing produced by oblique projection with hidden lines (d'), the former subject reported that it had a face on the floor (representing the bottom face of the cube) and might have side faces, which he could not interpret or understand. Another early blind subject, speaking about a drawing produced using oblique projection without hidden lines (d), reported similarly that the shape had one face (a square), but that she could not understand the other portions. Yet another early blind subject, referring to drawings by axonometric projection (e, e'), indicated that they had a parallelogram instead of a square, though he could not explain why. While these remarks suggest the subjects' understanding of the images drawn by these methods is not perfect, it seems that they are capable of understanding them at least partially, which might lead to eventual comprehension of the drawing methods.

4 Conclusion

The above results demonstrate that both early and late blind subjects had no difficulty understanding or producing drawings of geometric shapes used in the present study by both orthogonal projection and development. In addition, the results suggest that late blind subjects encountered little difficulty with respect to the oblique, axonometric, and perspective projection, at least when dealing with the cube, cuboid and cylinder, while early blind subjects had relatively significant difficulty. That being said, remarks made by early blind subjects during the course of procedure 3) seem to suggest that there is the possibility of their developing a greater understanding of the approaches of oblique, axonometric, and perspective projection, given additional exposure and educational processes with effective materials, for example ones mentioned in Introduction.

The above results also indicate that drawings of different geometric shapes produced via the same projection method have different degrees of ease of understanding for the blind, as shown in the rankings of drawings of the geometric shapes produced via the axonometric projection in procedure 2). Drawings of the cylinder via the axonometric projection method, compared to the ones of the cube or cuboid via the same method, seems to be easy to understand for early blind subjects. This finding seems to show that it is effective to consider such differences among various geometric shapes in the educational processes for understanding a certain projection method.

References

1. Tompson, L., Chronicle, E.: Beyond Visual Conventions: Rethinking the Design of Tactile Diagrams. *The British Journal of Visual Impairment* 24, 76–82 (2006)
2. Pathak, K., Pring, L.: Tactual Picture Recognition in Congenitally Blind and Sighted Children. *Applied Cognitive Psychology* 3, 337–350 (1989)
3. Mochizuki, T.: Visual and Tactual Depth Perception of 2-D Perspective Patterns in the Early or Congenitally Blind in the Case of Normal Sight [Abstract]. *Journal of Japan Women's University. Faculty of Home Economics* 26, 19–28 (1979)
4. Heller, M.A., Carcaterra, J.A., Burton, L.L., Tyler, L.A.: Tactual Picture Identification by Blind and Sighted People: Effects of Providing Categorical Information. *Perception & Psychophysics* 58, 310–323 (1996)
5. Heller, M.A., Brackett, D.D., Scroggs, E., Steffen, H., Heatherly, K., Salik, S.: Tangible pictures: Viewpoint effects and linear perspective in visually impaired people. *Perception* 31, 747–769 (2002)
6. Kennedy, J.M., Gabias, P., Nicholls, A.: Tactile Pictures. In: Heller, M.A., Schiff, W. (eds.) *The Psychology of Touch*, pp. 263–299. Lawrence Erlbaum, Hillsdale (1991)
7. Kennedy, J.M.: Drawings from Gaia, a blind girl. *Perception* 32, 321–340 (2003)
8. Kennedy, J.M., Juricevic, I.: Haptics and projection: Drawings by Tracy, a blind adult. *Perception* 32, 1059–1071 (2003)
9. Fujiyoshi, M., Kaneko, T., Fujiyoshi, A.: Development of Tactile Graphics Production Software for Three-Dimensional Projections. In: *Proceedings of ICCHP 2010. LNCS*, pp. 545–552. Springer, Heidelberg (2010)

Human-Computer Interaction and Usability Engineering for Elderly (HCI4AGING): Introduction to the Special Thematic Session

Andreas Holzinger¹, Martina Ziefle², and Carsten Röcker²

¹ Medical University Graz, A-8036 Graz, Austria
Institute for Medical Informatics, Statistics & Documentation (IMI)
Research Unit HCI4MED

andreas.holzinger@medunigraz.at

² RWTH Aachen University, D-52056 Aachen
Communication Science, Human Technology Centre (Humtec)
{ziefle,roecker}@humtec.rwth-aachen.de

Abstract. In most countries demographic developments tend towards more and more elderly people in single households. Improving the quality of life for elderly people is an emerging issue within our information society. Good user interfaces have tremendous implications for appropriate accessibility. Though, user interfaces should not only be easily accessible, they should also be useful, usable and most of all enjoyable and a benefit for people. Traditionally, Human-Computer Interaction (HCI) bridges Natural Sciences (Psychology) and Engineering (Informatics/Computer Science), whilst Usability Engineering (UE) is anchored in Software Technology and supports the actual implementation. Together, HCI and UE have a powerful potential to help towards making technology a little bit more accessible, useful, useable and enjoyable for everybody.

Keywords: Human-Computer Interaction, Usability Engineering, User Interfaces, Elderly People, Older Adults.

1 Introduction and Motivation for This Special Thematic Session

In most industrialized countries the demographical, structural and social trends tend towards more and more elderly people in single households, which definitely has effects on health care, emergency medical services and of course on the individuals themselves [1]. Older people and new technologies are one of the important research and development areas [2], where accessibility, usability and life long learning play a major role. For example, Intelligent User Interfaces (IUI) for Ambient Assisted Living (AAL) intend to support elderly by providing intuitive and natural interaction mechanisms [3]. However, such applications must be designed and developed to support the needs, the new, and special, demands and requirements of the individual end users. Clear benefits must be offered, whether in a physical, medical, emotional, motivational or educational respect.

The design and development of IT must support the elderly end users, especially to overcome their fears and enable them to accept technological aids and mobile devices without reservations. The design must then reflect the acceptance of the end users and not be the cause of new biases [4]. In order to ensure this, the HCI community developed a variety of User-Centered Design (UCD) techniques during the last 15 years [5], which are meanwhile established usability engineering methods [6].

2 Future Trends and Research Issues in HCI4AGING

The broad interpenetration and ubiquity of information and communication technologies have fundamentally changed the nature of social, economic and communicative pathways in modern societies [7][8]. Technology is present every-where and at any time and seems to overcome physical as well as mental borders. With the increasing diversity of users, the diversity of technical systems (visible vs. invisible, local vs. distributed) and the diversity using contexts (entertainment, medicine, office, mobility), new knowledge is needed for the deeper understanding of users' interaction with technology and technology acceptance [9]. If looking at current developments, a number of crucial factors come into fore that impact the barrier-free integration of future technologies:

User diversity: As opposed to the past, when mostly sophisticated and technology prone professionals were the typical users of technical products, increasingly older adults are confronted with a broad range of technology. Though, older users do have higher demands on usable interface designs [10]. Thus, one of the key issues of HCI designs is to understand the impact of older users' abilities and restrictions as well as their needs and wants respecting age-sensitive technical designs.

Function complexity and device miniaturization: The mobile character of technology in combination with the small communication window represents a still higher usability demand compared to large display technologies. The limited screen space is problematic for providing optimized information access, especially for aged users and those with low technical literacy [11]. In addition the increasing variety of functions capable of running on these devices imposes high cognitive load for users. In order to fully exploit the potential of (mobile) wireless technologies for a broad user group, research has to consider the impact of information designs and cognitive complexity.

New application fields: The application contexts, in which (mobile) technologies are used, changed considerably over time, and, still, application fields will tremendously change in future technology generations. Devices are used in different using contexts, in which technology has different roles, appearance, forms, and responsibility. This especially concerns smart home technologies and ambient intelligent environments [7], in which devices are invisibly communicating with remote computers, sometimes integrated in clothing [12], furniture or walls [9]. Thus, research efforts should address usability issues in complex technical environments.

Taken together, given these developments, usability requirements are more important than ever. As long as technical designs are not easy to use and learn, technical innovations will not have sustained success. Design approaches should therefore take

the user-perspective seriously. Technology must be developed in a way that users want to use it, and, still more important, that users are able to use it.

3 The 2010 Special Thematic Session Program

After a brief introduction to the special thematic session by the organizers, *André Calero Valdez* from RWTH Aachen University, Germany, presents the results of an exploratory study on mental models of mobile diabetes devices. Factors such as age, technology expertise, mobile phone expertise and medical domain knowledge were examined as mediating variables on the construction of mental models. The results show that the adequacy of the mental model and especially the lack thereof can heavily impact whether the initial usage of a device is successful or not. In this context, age as well as the lack of experience with mobile phones decreases the ability to form adequate mental models. In contrast, the users' domain knowledge did not significantly impact the quality of the mental model or the navigation performance.

This is followed by a talk of *Gemma Webster* from the University of Dundee, UK, who reports on the Portrait system. Portrait was designed for assisting medical care staff in learning important personal information about residents in nursing homes, thereby helping to stimulate conversations and discussions about the events outlined in the patients' profiles. The system consists of multimedia profiles of each resident presented on an easily and quickly accessible touch screen. Each profile contains information about the persons' key life events, family, preferences, as well as hobbies and interests. The talk closes by presenting the results of an initial usability study, carried out to evaluate the system with inexperienced computer users.

In the third talk, *Johannes Oberzaucher*, from the Living Lab Schwechat, Austria, reports on the development of a small and unobtrusive gait measurement system, which was especially designed for monitoring gait and body movements in unsupervised settings. The system can be used to record gait and body movement data in order to extract significant parameters for fall risk estimation. The developed insole system is used to determine spatiotemporal gait parameters during specified assessment procedures and thereby offers the opportunity to measure the bipedal gait and body movements wirelessly and directly at the foot in almost any surrounding while wearing normal walking shoes.

The fourth talk by *Jan Rennies*, from Fraunhofer IDMT, Oldenburg, Germany, addresses the design of a novel tool for quality monitoring of acoustic communication. The presented system is designed to create awareness for the particular problems of hearing-impaired listeners and to control further means for enhancing the communication quality. The presenter illustrates the technical development process as well as the experimental evaluation of the prototype system.

In the following talk, *Tomoyasu Komori* from the Science and Technical Research Laboratories of Tokyo, Japan, reports on the development of a device for evaluating broadcast background sound balance for elderly listeners. The system design was guided by the results gained in two investigative user studies. The first study analyzed the ability of elderly people to separate narration from background sound. In the second study, the authors explored the influence of the recruitment phenomenon and sound separation ability deterioration by analyzing the subjective evaluation results

and the program sound level. Based on these findings, an objective evaluation device was developed, which calculates the loudness of narration and background sound in real-time and indicates the appropriate sound balance for elderly people.

In the final talk of this session, *Guillaume Lepicard* from IRIT Toulouse, France, reports on an experimental user study analyzing touch-screen interfaces for older users. The study investigates the optimal position and number of tactile targets for touch-screen user interfaces intended for a diverse user population. The number of targets, the number of interaction blocks on the touch screen and the number of hands were used as independent variables in an experiment measuring realization time and accuracy. The presenter discusses the results and formulates design recommendations regarding touch-screen interaction for older people.

References

1. Kleinberger, T., Becker, M., Ras, E., Holzinger, A., Müller, P.: Ambient Intelligence in Assisted Living: Enable Elderly People to Handle Future Interfaces. In: Stephanidis, C. (ed.) UAHCI 2007 (Part II). LNCS, vol. 4555, pp. 103–112. Springer, Heidelberg (2007)
2. Emiliani, P.L., Stephanidis, C.: Universal access to ambient intelligence environments: Opportunities and challenges for people with disabilities. *IBM Systems Journal* 44(3), 605–619 (2005)
3. Mukasa, K.S., Holzinger, A., Karshmer, A.I. (eds.): Intelligent User Interfaces for Ambient Assisted Living. Fraunhofer IRB (2008)
4. Holzinger, A.: Usability Engineering for Software Developers. *Communications of the ACM* 48(1), 71–74 (2005)
5. Norman, D.A., Draper, S.: *User Centered System Design*. LEA, Hillsdale (1986)
6. Holzinger, A., Searle, G., Nischelwitzer, A.: On some Aspects of Improving Mobile Applications for the Elderly. In: Stephanidis, C. (ed.) HCI 2007. LNCS, vol. 4554, pp. 923–932. Springer, Heidelberg (2007)
7. Röcker, C.: Chances and Challenges of Intelligent Technologies in the Production and Retail Sector. *International Journal of Business, Economics, Finance and Management Sciences* 2(3), 150–161 (2010)
8. Ziefle, M., Jakobs, E.-M.: New challenges in Human Computer Interaction: Strategic Directions and Interdisciplinary Trends. In: 4th International Conference on Competitive Manufacturing Technologies, University of Stellenbosch, South Africa, pp. 389–398 (2010)
9. Ziefle, M., Röcker, C.: Acceptance of Pervasive Healthcare Systems: A comparison of different implementation concepts. In: 4th ICST Conference on Pervasive Computing Technologies for Healthcare 2010 (2010)
10. Arning, K., Ziefle, M.: Effects of Cognitive and Personal Factors on PDA Menu Navigation Performance. *Behaviour and Information Technology* 28(3), 251–268 (2009)
11. Ziefle, M.: Information presentation in small screen devices: The trade-off between visual density and menu foresight. *Applied Ergonomics*, 10.1016/j.apergo.2010.03.001 (2010)
12. Gaul, S., Ziefle, M.: Smart home technologies: Insights into generation-specific acceptance motives. In: Holzinger, A. (ed.) USAB 2009. LNCS, vol. 5889, pp. 312–332. Springer, Heidelberg (2009)

A Device to Evaluate Broadcast Background Sound Balance Using Loudness for Elderly Listeners

Tomoyasu Komori¹, Tohru Takagi¹, Koichi Kurozumi²,
Kiyotake Shoda³, and Kazuhiro Murakawa⁴

¹ NHK Science and Technology Research Laboratories 1-10-11 Kinuta,
Setagaya-ku, Tokyo, 157-8510 Japan

² NHK Engineering Service, INC, 1-10-11 Kinuta, Setagaya-ku, Tokyo, Japan

³ NHK Morioka 4-1-3 Ueda, Morioka-shi, Iwate, Japan

⁴ Yamaki Electric Corporation., Tokyo, 3-7-22 Shimomeguro Meguro-ku, Japan

Abstract. Elderly people complain that they sometimes have a hard time hearing the narration of broadcast TV programs because the background sounds (background music, sound effects) are too loud. We conducted subjective evaluations to determine the relationship between TV volume and loudness of background sounds as it regards the comprehension of programs by elderly subjects. On the basis of objectively measured loudness levels, we confirmed two conditions under which elderly listeners perceive background sounds as too loud. One is a less than 6 phon difference between the narration and background sounds; the other is when background sounds are more than 2.5 phon louder than the average narration in the program. Based on these findings, we constructed a prototype system for objectively evaluating loudness. The device features a meter with seven color-coded levels that clearly shows the best sound balance for elderly listeners. The evaluation system was tested at a broadcasting station.

Keywords: Loudness level, elderly listeners, broadcast program speech, subjective evaluation.

1 Introduction

Reports from a range of different sources indicate that elderly people often complain that the speech in TV programs is hard to hear, and they cite the reason to be that the background music and sounds are too loud. NHK has conducted a number of studies in an effort to make the speech in TV programs easier for elderly listeners to understand [1] [2] . It was found that in typical home listening environments, when the narration level stays the same and the background sound level fluctuates, elderly listeners' ability to hear is adversely affected by the recruitment phenomenon and noise in the house. Because of this, their ability to differentiate dialog from background sounds diminishes [3] [4] .

This problem led us to conduct a series of trials with program production mixers and to develop a system that uses the loudness level which the mixer determines is right for

the loudness of the background sounds (detailed in Section 3). We objectively evaluated the background sound loudness that fits the loudness derived by the mixers [5].

With the idea of developing a device for practical use, we conducted another round of subjective evaluations using actual broadcast programming. We found that elderly listeners did indeed perceive that the background sounds were too loud when the difference between the narration and the background sounds was small and when only background sounds were present but were louder than the average loudness of the narration. We incorporated these findings into our system for objectively evaluating the loudness of TV program background sounds and tested it at a broadcasting station. The results were encouraging, and they suggest that the system would be a valuable tool for producing programs with the optimum volume balance for elderly listeners.

2 Subjective Evaluations in Listening Booths

To gain insight into why elderly people have trouble hearing broadcast programs, we conducted a series of subjective evaluations in a listening booth environment.

2.1 Test Conditions

The video/audio material was an NHK documentary. The audio consisted of separate narration and background sound. A total of 49 minutes of content was divided into 45 segments of 30-150 s each, and each segment was treated as an evaluation set. Since the evaluation took a long time, we split the test period in half, with a short break in the middle. The background sound levels of the evaluation segments were adjusted to -6, 0, and +6 dB compare to the actual broadcast mixing level. We selected the background sound levels at random. The subjects evaluated the background sound loudness with respect to the narration on a five-level scale ranging from too loud to too soft (too loud, loud, fine, soft, too soft).

The trials were conducted on 12 elderly listeners ranging in age from 60 to 72. Some of the subjects had hearing loss in one ear, but we used the results from all 12 subjects in the subsequent analysis. The evaluation trials were done in a listening booth, and the content was presented using a digital audio workstation (DAW), a 22-inch LCD TV screen, and a pair of compact Entry-S (ALR-Jordan) speakers.

Two subjects were tested at the same time. During the trials, the narration was presented at three sound pressures (hereafter, listening levels) of 58, 65, and 75 dBA. These levels were thought to be in the approximate range of volume settings that elderly viewers might use when watching TV at home[3].

2.2 Trial Results

The narration listening levels of 58 and 65 dBA were chosen on the basis of the mixing balance derived from the reference level yielded by the program production mixer (engineer). On the other hand, the assessments of the background sounds as "loud" or "too loud" went up by 25% when the narration was at the 75 dBA, although they did not present any hearing problems. In other words, elderly listeners perceived the background sounds as getting louder the listening level increased, even though the mixing balance was the same.

On the other hand, we found cases in which the mixing balance that was judged appropriate by the program production mixer was evaluated to be inappropriate by the elderly listeners and their response depended on production effects and even the evaluation set segments. For example, we found that if the loudness level of the sound effects increased, the psychological effect also increased, even at the reference level. In other words, the volume of the reference level background sound was adjusted dynamically from segment to segment and even within segments in order to give a intended production effects.

These results led us to convert the sound source levels into loudness levels and to examine the evaluation results in order to isolate the conditions under which elderly listeners perceive background sounds as getting louder.

3 Analysis Based on Loudness Level

Measuring loudness levels during very short time intervals—that is, measuring loudness or loudness levels in real time—is an effective way to assess dynamic changes in the loudness of narration and background sounds. [6]

3.1 Loudness Levels Used in the Analysis

The calculations were done based on ISO 532B [7] and ISO 226 (2003) [8] compliant methods [9] [10] . In these methods, speech signals are partitioned into frames of short duration (about 42 ms per frame), with 50% overlap (21 ms shift width), and the instantaneous loudness level is calculated for each frame. However, it has been shown that program sound loudness levels depend entirely on the larger values among the instantaneous loudness levels. More specifically, the program sound loudness level is defined as "the average of the largest 4 values in the 2 s immediately preceding any given time taken as standard within the instantaneous loudness level" [11].

Moreover, we employed a "weighted loudness level" using zero frequency weighting when calculating loudness levels for frequencies below 250 Hz [12]. Subsequent references to loudness level shall refer to this weighted loudness level.

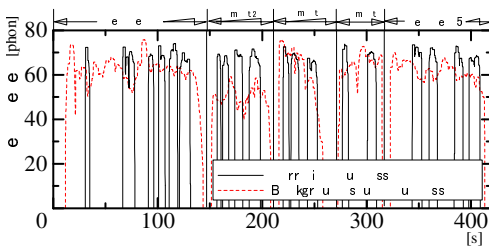


Fig. 1. The calculated narration and background sound loudness levels (There is 10 s time for response every end of segment)

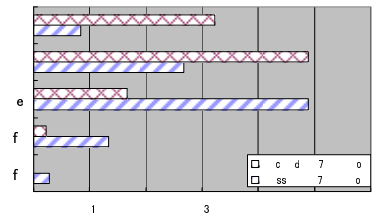


Fig. 2. Histogram of evaluation comparing different background sound loudness levels (loudness levels for background sounds alone)

Fig. 1 shows the calculated narration and background sound loudness levels for five segments out of the evaluation set from the beginning of the program. While the narration loudness level remains approximately constant, one can see that the background sound loudness level varies from segment to segment and even fluctuates from moment to moment within segments.

3.2 Effect of Playing Background Sound Alone

We know from the previous work of Kurozumi et al. [11] that a difference in loudness of 2.5 phon or greater will be discernible to an ordinary listener. A narration listening level of 75 dBA is equivalent to a narration loudness level of 70 phon. Therefore, we predicted that a segment would annoy elderly listeners if the background sound of the evaluation set segments exceeded 72.5 phon. We divided the evaluation results into two groups and examined them. The two groups' background sounds were respectively more or less than 2.5phon louder than the average narration loudness level.

Fig. 2 is a histogram of the evaluations of all 12 subjects of the background sound evaluation segments exceeding or less than 72.5 phon. One can see that the elderly listeners perceived the background sounds with segments exceeding 72.5 phon as louder than background sounds without segments exceeding 72.5 phon.

3.3 Effect of Background Sound on Narration

We conducted the experiment with program production mixers using the programming content evaluated by the subjects. The experiment result showed that the loudness level of the background sounds of the documentary was reduced to 6-12 phon compared with the loudness level of the narration [5]. It is thought that the program production mixers would determine that background sounds of 6 phon is just barely loud enough for broadcast programming.

To test the prediction that elderly listeners would perceive the background sounds as too loud when the difference in loudness levels between the narration and background sound is less than 6 phon, we examined the distribution of the evaluation results for the set segments in which the difference between narration and background sound was less than 6 phon or greater than 6 phon. Assuming that the ability to hear the narration would not be affected by loud background sounds of short duration, we set the time threshold for the less than 6 phon difference at 4 s.

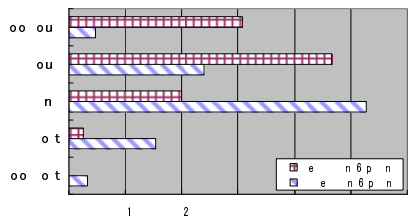


Fig. 3. Histogram of evaluations when background sound loudness levels were 6 phon more than or less than that of the narration

Fig. 3 is a histogram of the evaluation results for all 12 subjects of the evaluation set segments with loudness level differences of less than 6 phon or greater than 6 phon.

One can see that the elderly listeners perceive the background sounds as louder when the difference between sound levels is less than 6 phon compared with when the difference is greater than 6 phon.

3.4 Analysis of Variance Results

We conducted a two-way analysis of variance (ANOVA) test on all 45 evaluation segments. We divided the results into four groups depending on (a) whether background sounds were more or less than 72.5 phon, for replayed background sounds alone, and (b) whether the loudness level difference between narration and background sounds was less than or more than 6 phon. It was found that there is a significant difference ($p < 0.01$) between background sound above and below 72.5 phon and between narration and background sound differences of more than or less than 6 phon. However, no significant difference regarding an interaction between factors was found.

3.5 Implications and Considerations

One of the causes that make elderly listeners tend to perceive background sound in excess of 72.5 phon as too loud is the recruitment phenomenon, whereby listeners have difficulty hearing soft sounds but louder sounds seem too loud. Six of our subjects tested positively for the recruitment phenomenon [4], but even other six subjects with normal hearing are likely to perceive background sounds as too loud at listening levels above 75 dBA, even when program production mixers indicate that the sound level is right.

4 Prototype Evaluation System

We constructed a prototype objective evaluation system that incorporated the findings described above. The system calculates narration and background sound loudness in real time, and displays the background sound balance optimized for elderly viewers on a seven-level color-coded meter. The system also features a signal processing function enabling the signal to be simulated so that people with normal hearing can experience how elderly listeners perceive.

4.1 Objective Evaluation Using the System

Fig. 4 and Fig. 5 show a block diagram and a photograph of the system. Separate stereo narration and background sound input to the calculation unit, and fast Fourier transform (FFT/IFFT) processing is performed on 2,048 points on four channels for every shift of approximately 21 ms. The loudness levels of the narration and background sounds are calculated from the audio signals that the FFT converted into the frequency domain, and the results are sent to the display unit through a USB cable. At the same time, signal processing is done to simulate the hearing of elderly listeners.

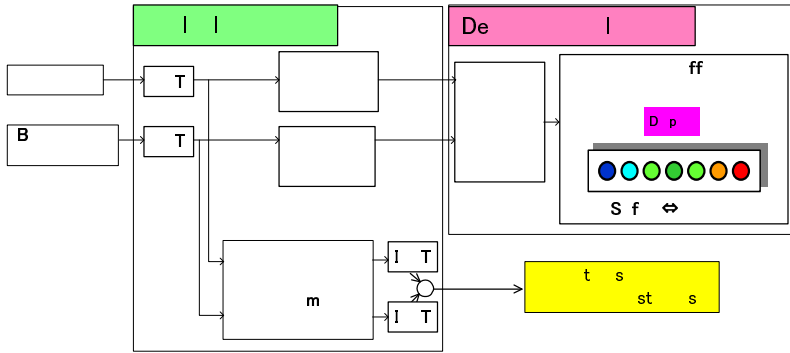


Fig. 4. Block diagram of prototype evaluation system



Fig. 5. Prototype evaluation system (The display unit sits on top of the calculation unit)

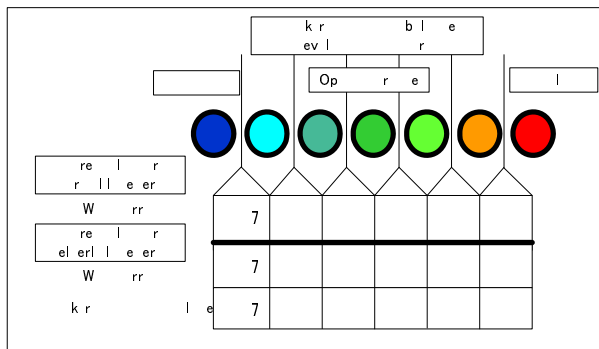


Fig. 6. Loudness level thresholds when listening level is 70 phn

The background sound balance is represented a series of thresholds based on the loudness level, the difference in loudness level between the narration and background sound, and the reference listening level, and it is displayed as a seven color-coded scale (blue, aqua, blue-green, green, yellow-green, orange, red), with blue on the left signifying too soft, red on the right signifying too loud, and green in the middle signifying the optimum sound level. Fig. 6 shows the threshold values for a listening level of 70 phn. The figure compares background sound balance thresholds for people

with normal hearing as determined with our previous mixer [5] and thresholds for elderly listeners.

The prototype evaluation system is also equipped with signal processing capabilities that manipulate each auditory band to simulate a diminished capacity of differentiating speech from background sounds and the recruitment phenomenon. These capabilities enable program producers to experience firsthand how elderly listeners are hearing in their programs [2].

4.2 Field Test of Prototype System at a Broadcasting Station

The prototype evaluation system was deployed at the Morioka broadcasting station in January 2010, and the feedback assessing its performance was very favorable. The station personnel were especially pleased with the easy-to-understand display. They found that even increasing the volume by a small degree when only background sounds were present caused elderly listeners to complain that the background music and sounds were "noisy." This sensitivity of elderly listeners to moderate increases in volume came as a surprise to them. Moreover, the personnel gained new insight into how elderly listeners perceive broadcast programming by listening to the system simulating sound as heard by typical seniors. We are now incorporating a number of refinements and improvements suggested in feedback from the station personnel into the prototype system.

5 Conclusions

We confirmed that elderly listeners perceive that sound in TV programs is too loud (a) when the difference between narration and background sounds is less than 6 phon and (b) when background sounds are more than 2.5 phon louder than the average narration level when only background music or sounds are present. On the basis of these findings, we developed a prototype system for objectively evaluating the loudness of broadcast TV programs optimized for elderly listeners [2]. The prototype system was tested at an actual broadcasting station and found to be an extremely useful tool for aiding in the production of TV programs with the right sound volume balance for elderly listeners. We will subject the new system to a thorough evaluation process and to make sure it does what it was designed to do: reveal the proper sound balance for elder viewers.

The programming content used in this study was a documentary in which the narrator's voice is very clear. However, there is a wide spectrum of programming—dramas, quiz shows, and so on—in which the speech is not so clear. We would like to conduct another survey that includes the "clarity of speech" as a variable affecting the ability of elderly listeners to hear and understand.

References

1. Nakamura, H., Sawaguchi, M., Masaoka, K., Watanabe, K., Yamasaki, Y., Miyasaka, E., Yasuoka, M., Seki, H.: Better Audio Balance Broadcasting service for elderly people.- Back Ground Sound Levels of Television Programs for Easy Listening. In: 2003 Proc. Spring Meet. Acoust. Soc. Jpn.,1-5-5, pp. 455–456 (2003) (in Japanese)

2. Komori, T., Takagi, T., Kurozumi, K., Murakawa, K.: An Investigation of Audio Balance for Elderly Listeners using Loudness as the Main Parameter. AES 125th Convention paper 7629 (2008)
3. Komori, T., Takagi, T.: Subjective evaluation on the background sound of TV program by elderly people in the home environment. In: Proc. Autumn Meet. Acoust. Soc. Jpn., 3-10-14 (September 2009) (in Japanese)
4. Komori, T., Takagi, T.: A study of elderly people's hearing loss and the subjective evaluation result from varying background sound level of TV program. In: ITE Winter Annual Convention, 4-9 (2009) (in Japanese)
5. Komori, T., Komiyama, S., Dan, H., Takagi, T., Shoda, K., Kurozumi, K., Hoshi, H., Murakawa, K.: A Investigation of the Audio Balance Control based on the Loudness Level. IEICE Transactions J92-A(5), 344–352 (2009) (in Japanese)
6. Kurozumi, K., Okamoto, N.: Real-time Loudness Level Meter for Broadcasting Stations and Recent Trends of Standardization. Journal of the Institute of Image Information and Television Engineers 60(7), 1012–1017 (2006) (in Japanese)
7. ISO532, Acoustics - Methods for calculating loudness level (1975)
8. ISO226, Acoustics - Normal equal loudness level contours (2003)
9. Paulus, E., Zwicker, E.: Computer programs for calculating loudness from 1/3 octave band levels or from critical band levels. Acoustica 27, 253–266 (1972)
10. Zwicker, E., Fastl, H., Widmann, U., Kurakata, K., Kuwano, S., Namba, S.: Program for calculating loudness according to DIN 45631 (ISO 532B). J. Acoust. Soc. Ja(E) 12, 39–42 (1991)
11. Kurozumi, K.: Loudness level meter based on the ISO226(2003-08-15). In: Proc. Autumn Meet. Acoust. Soc. Jpn., 1-5-10, September 2004, pp. 495–496 (2004) (in Japanese)
12. Miura, T.: Auditory Perception and Speech. I.E.I.C.E., 448–453 (1980) (in Japanese)

Automatic Live Monitoring of Communication Quality for Normal-Hearing and Hearing-Impaired Listeners

Jan Rennies, Eugen Albertin, Stefan Goetze, and Jens-E. Appell

Fraunhofer IDMT, Hearing, Speech and Audio Technology, Oldenburg, Germany
jan.rennies@idmt.fraunhofer.de

http://www.idmt.fraunhofer.de/eng/hearing_speech_audio_technology/index.htm

Abstract. This contribution presents a system, which allows for a continuous monitoring of speech intelligibility from a single microphone signal. The system accounts for the detrimental effects of environmental noise and reverberation by estimating the two relevant parameters signal-to-noise ratio and reverberation time, and feeding them to a speech intelligibility model. Due to its real-time functionality and the fact that no reference signal is required, the system offers a wide range of opportunities to monitor communication channels and control further signal enhancement mechanisms. A priori knowledge of the individual hearing loss can be used to make the system applicable also for hearing-impaired users.

1 Introduction

At modern workplaces as well as in the private sphere, chat or video-conferencing systems and mobile devices have become an integral part for acoustic communication. Under adverse acoustic conditions, however, factors such as background noise, competing speech or reverberation can reduce the quality of the communication or even prevent the user of such systems from understanding at all. Particularly for hearing-impaired people, non-optimal acoustic conditions can considerably influence the ability to communicate. In contrast to visual impairment, which can be cured by purely passive devices such as lenses or glasses in most cases, even the most recent hearing aids cannot completely restore the functionality of the complex hearing system. Further technical support in acoustic communication systems for hearing-impaired people is therefore highly desirable.

In modern teleconferencing systems, one problem of hearing-impaired users and also often of normal-hearing people is that the near-end speaker is not aware of the potentially low intelligibility at the location of the distant listener. Without technical support the distant listener needs to make constant interventions, which obviously disturbs the communication. This contribution describes a novel tool to automatically monitor the quality of acoustic communication, which can be used to create awareness for the particular problems of the far-end listener and to control further means to enhance the communication quality, e.g. by

acoustic signal processing schemes for noise reduction [1,2] or dereverberation [3]. Estimating the relevant physical parameters from the signal recorded by a microphone, the system makes use of a well-evaluated speech intelligibility model [4,5].

The remainder of this contribution is organized as follows. Section 2 provides a detailed description of the implementation of the system and its components as well as an extension for hearing-impaired listeners, before the system is evaluated in Section 3. Section 4 concludes the paper.

2 Extended Speech Intelligibility Model for Real-Time Monitoring of Acoustic Communication

The schematic structure of the system is shown in Fig. 1. Let k denote the discrete time index. The acoustic signal $y[k]$ recorded by the microphone can be described as a mixture of the speech signal $s[k]$ and environmental noise $n[k]$, i.e. $y[k] = s[k] + n[k]$. The speech is affected by the acoustic properties of the room, which can mathematically be described by the so-called room impulse response (RIR) $h[k]$. The reverberant speech signal $s[k]$ is the convolution of the clean speech signal $s'[k]$ and the RIR $h[k]$, i.e. $s[k] = s'[k] * h[k]$. For practical applications, the system needs to work with low delay. The recorded signal is therefore analyzed in blocks of length L . For each block ℓ , signal processing strategies are used to extract the main physical parameters affecting speech intelligibility, namely signal-to-noise ratio (SNR) and reverberation time (T_{60}) (see Section 2.1). These two parameters are averaged for each block and subsequently processed by a speech intelligibility model, which relates the physical parameters to a perceptual quantity. The basic principle of the speech intelligibility model is the same as used in the well-known Speech Transmission Index (STI) [4,5], which is widely applied both in research and practical systems. It is beyond the scope of this contribution to detail on the concept underlying the STI, the interested reader is referred to the literature. Briefly, the effects of reverberation and background noise are characterized by the modulation transfer function (MTF), which is related to perceived intelligibility after accounting for (among other things) auditory masking, the different importance of certain frequency regions for speech recognition and the threshold of hearing. For hearing-impaired listeners, the system can be provided with information from the audiogram (see

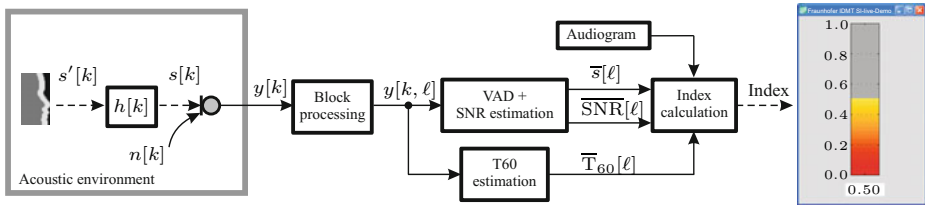


Fig. 1. Schematic structure of the processing stages for the system to estimate speech intelligibility from a single microphone signal

Section 2.2). The calculation of the STI results in an index between 0 and 1, where 0 indicates no intelligibility and 1 indicates perfect intelligibility. Intermediate values of the index can be transformed into other measures for intelligibility like e.g. percentage of correctly understood words or sentences (see for example [6]). At the output of the system, an intelligibility index is available for each block, and can be graphically displayed or used for further processing.

2.1 Estimation of Physical Parameters

As mentioned above, signal-to-noise ratio (SNR) and reverberation time (T_{60}) are the main factors affecting speech intelligibility. For estimation of the SNR the system uses a single-channel noise-reduction technique [1] combined with a voice activity detector (VAD) [2]. This technique provides an estimation of the SNR in several frequency bands n for each short-time frame ℓ . The SNR estimation is used for those frames when voice activity was detected.

The estimation of reverberation time (T_{60}) is based on the cepstral averaging procedure [7], which uses the theory of blind homomorphic deconvolution to convert the convolution $s[k, \ell] = s'[k, \ell] * h[k, \ell]$ into a sum of cepstra. If n denotes the discrete frequency index, \Re denotes the real part and $\text{IDFT}\{\cdot\}$ denotes the discrete inverse Fourier transform, the equivalent description of the convolution in the frequency domain is $S[n, \ell] = S'[n, \ell] \cdot H[n, \ell]$, and the cepstra can be calculated as

$$c_s[k, \ell] = \Re \{ \text{IDFT} \{ \log(S'[n, \ell]) + \log(H[n, \ell]) \} \} = c_{s'}[k, \ell] + c_h[k, \ell] \quad , \quad (1)$$

where $c_{s'}[k, \ell]$ and $c_h[k, \ell]$ are the cepstra of the clean speech and the RIR, respectively. Under the assumption that $s'[k]$ is not strongly correlated for different blocks ℓ , averaging over several blocks N gives an estimate of the cepstrum of the RIR

$$\hat{c}_s[k] = \sum_{i=1}^N c_{s'}[k, i] + \sum_{i=1}^N c_h[k, i] \approx \hat{c}_h[k] \quad . \quad (2)$$

Inverting the cepstrum as $\hat{h}[k] = \text{IDFT} \{ \exp(\text{DFT} \{ \hat{c}_h[k] \}) \}$ yields an estimate of the RIR, from which the reverberation time can be calculated (for details, see [7]).

2.2 Effects of Hearing Impairment

Apart from physical parameters, the calculation of the STI also uses quantities related to the auditory system (see [5] for details). One of these quantities is the threshold of hearing, i.e. the minimum level at which sounds are audible. In typical applications of the STI, normal hearing is assumed. However, since one of the main indications of hearing impairment is an elevated hearing threshold, it is possible to modify the threshold of hearing in the system to account for hearing losses. Obviously, information about the individual hearing loss of the listener cannot be estimated from the acoustic signal, and therefore has to be provided

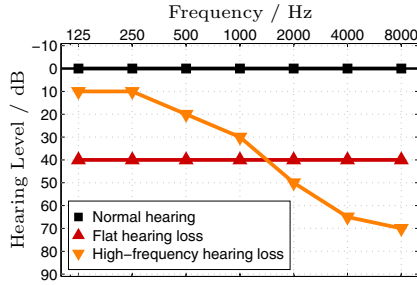


Fig. 2. Audiogram for normal-hearing (squares) and a person with flat (upward-pointing triangles) and high-frequency hearing loss (downward-pointing triangles)

as a priori knowledge. The most common measure for hearing impairment is the audiogram, which indicates the individual hearing threshold relative to the average normal-hearing person (see Fig. 2). The audiogram shows the threshold of pure tones (expressed in dB Hearing Level) as a function of the frequency of the pure tone. A value of zero indicates normal hearing. For clinical applications, a hearing level of up to 15 to 20 dB is considered normal, higher levels indicate a hearing loss. Exemplary audiograms of a normal-hearing (squares) and two hearing-impaired persons (triangles) are shown in Fig. 2. The hearing loss indicated by upward-pointing triangles is a flat hearing loss, i.e. the hearing level is similar for all frequencies. Such a hearing loss is usually observed e.g. when the middle ear is damaged. The hearing loss indicated by downward-pointing triangles is a high-frequency hearing loss, i.e. the hearing levels increase with frequency. Such a shape is typical for age-related hearing losses. The system presented here can include the effects of hearing impairment in terms of an elevated hearing threshold as indicated by the audiogram (see Fig. 1). Such an extension of the STI was already presented in [8].

In the following section, the performance of the proposed system is evaluated with respect to physical changes of the acoustic signal as well as to the effects of hearing impairment.

3 Evaluation of the Influence of SNR, Reverberation and Hearing Impairment

3.1 Methods

To evaluate our system, we tested it using continuous speech from a well-established speech test (Oldenburg sentence test, [9]). The speech was generated by concatenating sentences of the speech test in random order without pauses between the sentences. The speech was subsequently superimposed by a stationary noise, which had the same long-term spectrum as the speech. The signals were reverberated by convolution with an RIR to mimic the room acoustics of a typical office room ($T_{60} = 300$ ms). The intelligibility index was estimated every

2 s. Initially, the speech was superimposed by the stationary noise of relatively low level at an SNR of 15 dB, resulting in good speech intelligibility. After 20 s, the SNR was lowered to 10 dB. In practice, such a reduction in SNR could occur when an additional noise source is switched on. After 40 s, the SNR was further decreased to 0 dB. After 60 and 80 s, the SNR was increased to 5 and 15 dB, respectively, corresponding to a step-wise reduction of the noise source. The intelligibility index calculated with our system for this scenario of time-varying acoustic conditions is shown in Fig. 3 for a period of 100 s. The time instances when the changes in SNR occurred are indicated by vertical dashed lines. To investigate the sensitivity of the system to SNR, reverberation and hearing impairment, the calculations were made based on (i) SNR estimation alone (squares), (ii) based on SNR and T_{60} (circles) and (iii) based on SNR and hearing impairment (triangles) using the same audiograms (and symbols) as in Fig. 2. In condition (i) and (iii), the system assumed no reverberation, i.e. a totally anechoic room.

For all estimation methods, the effect of SNR can be clearly observed: the predicted intelligibility index is highest in the beginning and in the end, when the SNR is 15 dB. Accordingly, it is lowest for an SNR of 0 dB and at intermediate values for SNRs of 5 and 10 dB. The data also indicate slight variations in the predicted index, even when the SNR is constant. This can be explained by two effects. On the one hand, the energy distribution of running speech varies over time, which leads to a time-varying SNR. On the other hand, the estimations of both SNR and reverberation time are based on statistical methods [1,2], which introduce fluctuations in the prediction. In particular at higher SNR, the estimated index is lower, when the reverberation time is considered for calculation of the intelligibility index. This reflects the stronger detrimental effect of reverberation on speech intelligibility at higher SNRs, whereas intelligibility is primarily decreased by background noise at a low SNR. The two exemplary hearing losses included in the analysis result in a decreased intelligibility index for all SNR, which is expected since part of the speech energy is lower than the threshold of the assumed hearing losses.

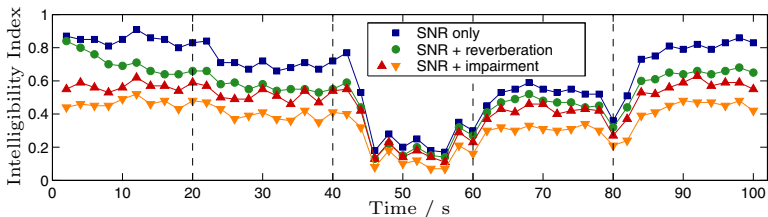


Fig. 3. Estimated intelligibility index as a function of time calculated by the monitoring system. The calculations are based on estimations of both SNR and T_{60} (circles) or SNR alone (squares). Dashed lines indicate points of time at which the SNR was changed (see text). Triangles represent estimations including hearing impairment as indicated in Fig. 2.

3.2 Results

The dependence of the intelligibility index estimated by the proposed system on SNR, reverberation time and hearing impairment shown in Fig. 3 indicates that the system can qualitatively predict the influence of these parameters on speech intelligibility. To verify whether it is also possible to derive a quantitative prediction, the estimated intelligibility was compared to an index that was computed based on a priori knowledge of SNR and T_{60} , i.e. the same index was also computed without estimating these parameters. The correlation between the calculated intelligibility index which is based on perfect knowledge of SNR and T_{60} and the estimated index is shown in Fig. 4 for the same 100 s of speech in noise and the same four conditions as in Fig. 3. Squares indicate index pairs based on an estimation (or a priori knowledge) of the SNR only, circles represent data derived from both SNR and T_{60} , and triangles indicate data including hearing impairment. Ideally, the values of estimated and calculated intelligibility index should be the same. In this case, the system would deliver the same predictions as the well-evaluated STI. Figure 4 shows that this is not the case. In general, the calculated index is higher than the estimated index. However, the deviation is systematic, and the linear correlation between estimated and calculated index is highly significant (correlation coefficient $R=0.86$, p -value <0.001). Further comparison revealed that the index pairs that correspond to time instances when the SNR changed are of particular interest. These data points are highlighted as gray symbols with black edges in Fig. 4. While most of the index pairs fit into the linear relation between calculation and estimation, some data points deviate from this trend, and it can be seen in Fig. 4 that these data points correspond to the index values at the time instances when the SNR was changed. The reason is that the system averages the estimated values over the duration of one block, and in consequence the output of the system is delayed by one block compared to the calculated index. At sudden changes of the SNR, the real intelligibility changes faster than the system can follow, resulting in one sample of over- or underestimated intelligibility. In practical applications, this delay may be a

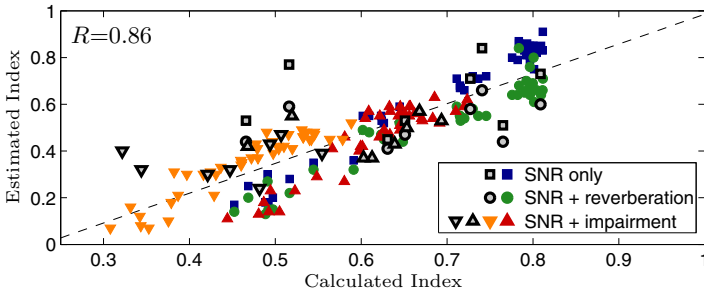


Fig. 4. Correlation between estimated and calculated intelligibility index. The symbol shapes are the same as in Fig. 3 (see text in Section 3). The dashed line indicates the linear regression $y = 1.276 \cdot x - 0.261$ obtained using a least-squares fit.

minor problem, since such sudden changes in SNR do not usually occur. Neglecting those points in the correlation analysis leads to $R=0.90$.

4 Summary and Conclusions

The described intelligibility monitoring system model is sensitive to changes in the acoustic environment such as ambient noise and reverberation as well as to different degrees of hearing impairment. Using the acoustic signal of a single microphone, the perturbing influence of room acoustic properties on speech intelligibility is accounted for by estimating reverberation time and SNR. These estimations are inherently subject to uncertainties. In general, the longer the block over which the parameters are averaged, the more accurate the estimations become. For practical use, a reasonable trade-off between estimation accuracy and the ability to react to changes in the acoustic environment requiring shorter averaging times has to be adjusted. This trade-off is application specific. Here we used an update interval of 2 s. This may be appropriate for telecommunication scenarios in which the near-end talker shall be provided with information on the intelligibility of the far-end listener. However, in case our system is used to control algorithms for acoustic signal processing, other settings may be required.

Despite the fact that the system has a slightly delayed response to sudden changes of the acoustic conditions, the linear correlation between the estimated intelligibility index and the same index calculated using a priori knowledge of the relevant parameters is highly significant. This makes the systems in principle applicable for the online monitoring of intelligibility in real applications. This constitutes a major advantage of the system compared to well-established speech intelligibility models (e.g. [5,10]), which are typically based on long-term spectra and require speech and noise signals to be available separately (i.e. they require a priori knowledge of the SNR).

The extension of the system to include information of the audiogram represents a first step towards a monitoring system also applicable for hearing-impaired listeners. Current research investigates the effects of other aspects of hearing impairment on speech intelligibility such as for example reduced frequency selectivity, reduced temporal resolution or modified loudness perception. In principle, models describing these effects could be integrated in the presented system to achieve more accurate predictions for hearing-impaired listeners.

Together with research for more elaborate models for hearing impairment, future research will aim at an increased estimation accuracy of the system's components by comparing several methods established in the literature, e.g. blind T_{60} estimation by means of autocorrelation analysis [7] or SNR estimators for fluctuating noise sources [2,11]. In order to optimize the system, further evaluations comprising different speech and noise signals, room acoustic conditions and degrees of hearing impairment as well as evaluations involving normal-hearing and hearing-impaired listeners will be made.

Due to its real-time functionality and the fact that no reference signal is required, the system offers a wide range of opportunities to monitor communication channels, e.g. direct feedback to the speaker about the particular problems

of the listener, the control of further signal enhancement mechanisms like noise reduction or dereverberation, or the control of output modalities in modern user interfaces.

References

1. Ephraim, Y., Malah, D.: Speech enhancement using a minimum mean-square error log-spectral amplitude estimator. *IEEE Transactions on Acoustics, Speech and Signal Processing* 33(2), 443–445 (1985)
2. Marzinzik, M., Kollmeier, B.: Speech pause detection for noise spectrum estimation by tracking power envelope dynamics. *IEEE Transactions on Speech and Audio Processing* 10(2), 109–118 (2002)
3. Habets, E.: Single and Multi-Microphone Speech Dereverberation using Spectral Enhancement. PhD thesis, University of Eindhoven, Eindhoven, The Netherlands (June 2007)
4. Houtgast, T., Steeneken, H.: A review of the MTF concept in room acoustics and its use for estimating speech intelligibility in auditoria. *J. Acoust. Soc. Am.* 77, 1069–1077 (1985)
5. International Electrotechnical Commission: IEC 60268-16 Sound System Equipment - Part 16: Objective rating of speech intelligibility by speech transmission index (1998)
6. Fletcher, H., Galt, R.: The perception of speech and its relation to telephony. *J. Acoust. Soc. Am.* 22, 89–151 (1950)
7. Schröder, J., Rohdenburg, T., Hohmann, V., Ewert, S.D.: Classification of reverberant acoustic situations. In: Boone, M. (ed.) *Proceedings of the International Conference on Acoustics NAG/DAGA 2009*, pp. 606–609. DEGA e.V, Berlin (2009)
8. Holube, I., Kollmeier, B.: Speech intelligibility prediction in hearing-impaired listeners based on a psychoacoustically motivated perception model. *J. Acoust. Soc. Am.* 100, 1703–1716 (1996)
9. Wagener, K., Brand, T., Kollmeier, B.: Development and Evaluation of a German Sentence Test. In: *Contributions to Psychological Acoustics - 8th Oldenburg Symposium on Psychological Acoustics*, pp. 439–466. bis-Verlag, Oldenburg (2000)
10. American National Standards Institute: ANSI S3.5-1997 Methods for calculation of the speech intelligibility index (1997)
11. Tchorz, J., Kollmeier, B.: SNR estimation based on amplitude modulation analysis with applications to noise suppression. *IEEE Transactions on Speech and Audio Processing* 11(3), 184–192 (2003)

Portrait: Portraying Individuality

Gemma Webster, Deborah I. Fels, Gary Gowans, and Norman Alm

School Of Computing and Duncan of Jordanstone College of Art and Design,
University of Dundee, Dundee, Scotland, DD1 4HN
{gwebster, debfels, nalm}@computing.dundee.ac.uk,
g.m.gowans@dundee.ac.uk

Abstract. People with dementia who live in care homes can have very little social interaction. Care staff have limited time to spend with each person and communication difficulties can make it difficult to get to know the person with dementia as a *person*. This paper presents Portrait a software tool to enable care staff to get to know a person with dementia quickly. An initial usability study was carried out to evaluate the system with inexperienced computer users. The study was conducted in two iterations collecting data on ease of use, preference of features, level of training required and how engaging Portrait was to use. Overall Portrait was very positively received with no major usability issues and all participants rated the system as either engaging or very engaging and fun to use.

Keywords: Multimedia, Dementia, Personality, Person Centred, Care Aid.

1 Introduction

Dementia is an umbrella term used to describe intellectual impairments normally in older people produced by brain dysfunctions that are usually progressive and eventually severe. People with late-stage dementia often start to lose the ability to communicate verbally and this means they can be very difficult to get to know. This is particularly problematic for people introduced into care homes with late stage dementia, which can often be the case as a result of families being no longer able to manage at home. These individuals will be unknown to the care providers and less able to communicate who they are and what they need. As a result, dementia care can follow a pattern of only dealing with the clinical care needs of a person as prescribed by their care plan and daily living activities such as eating, washing and toileting [1] with little social interaction identified as important to the health and well-being of a person with dementia [2, 3, 4].

Past life experiences and a person's history can help care staff to interact with residents and see beyond the disease [5]. It is common for care homes and families to create a life story, memory or scrap book for the person in care to be used to help care staff interact with the person with dementia through conversation and reminiscence [6]. While these books are very useful and helpful in some ways, it can be difficult for care staff to use them independently to better understand a person in a quick and informative manner as these books are often large binders containing detailed

information about the person's life. In addition, they do not point staff to the most salient information about a person and translate that into improved quality of care.

Tools created to aid people with dementia can vary from very low-tech pill boxes to sophisticated multimedia biography systems. One example is the Multimedia Biographies project which is designed to tell the life story of a person with Alzheimer's disease and mild cognitive impairments using family photos and videos, audio narration and music examples from that person's life and organized on a DVD [7, 8]. A different approach is taken by the Computer Interactive Reminiscence and Conversation Aid (CIRCA) project which presents generic multimedia content consisting of images, videos and music using a touch-screen interface. CIRCA is designed to be used by a person with dementia and their carer as a conversation aid [9, 10]. These example tools are designed as communication or reminiscence aids to be used along with the person with dementia but do little to help non-family caregivers discover the most important elements of a person's life or preferences prior to these interactions. In addition, these types of systems may not be available in every institution.

In this paper, a new system called Portrait, will be presented along with initial usability results for novice users. The system was devised to assist care staff, who are often inexperienced computer users, in learning some of the most important personal information about a resident.

2 System Design

The Portrait system consists of multimedia profiles of each resident presented on an easily and quickly accessible touch screen interface. Each profile contains information about the person's key life events, family, important things to know, preferences, and hobbies and interests. It is designed so that care staff can use the system alone in a setting such as a work room and find interesting information about a person in approximately three minutes. It is *not* intended to be used as an electronic story book or to be used with the individual being profiled. However, it may assist care staff in stimulating conversation and discussions with the person about the events outlined in his/her profile.

As the target user group was identified as inexperienced users with a large variation in literacy levels and age, it was important to design the interface to be simple and easy to use [11]. Little or no training should be required to access the profile information in the system.

Albinsson & Zhai (2003) suggest that touch screen interaction is one of the most direct and simplest forms of HCI as the information display and control are one surface. In addition, touch screens are widely used in public spaces and accepted for applications such as banks, museums, as ticketing machines and information kiosks [12, 13]. As a result we have implemented a touch screen interface for Portrait.

Although there is a growing body of research on how to design interfaces for older people and people with dementia, there is little published on the interaction of care staff with computers. Care staff have very little, if any, need to interact with a computer in their normal working tasks. However, studies have shown that information

retention and engagement are improved for people when interactive multimedia is used particularly in learning environments [14]. Using a multimedia system to deliver care information could have many benefits over a standard paper equivalent for care staff.

As a result, Portrait is designed as a multimedia interactive system. The system uses icons and text for main headings and limited text descriptions in the form of short concise sentences (see Figure 1 & 2). Images are used to allow care staff to quickly identify the topic or event in the system with text supporting or explaining the image in slightly more detail. Sound has not been included in Portrait as care environments are often noisy and it could be difficult for care staff to hear the sound. Portrait may also not be in a separate location and sound could be very distracting for residents or other care staff that are carrying out other tasks. Colour is used in the interface design to help users identify where they are within Portrait. The menu option and background colour of a topic (see Figure 1 & 2) are the same to help users identify where they are in the interface. To ensure the system was not too confusing or time consuming there are a maximum of two levels. Level 1 is opened when the user selects a menu option for a specific topic from the main screen. Upon selection the information for that topic is displayed (see Figure 1) and the user has the option to progress to Level 2 to view additional information (see Figure 2). There is a back button for Level 2 allowing the user to return to Level 1. However, the user can select a different topic at any time from the main topic menu consistently located at the bottom of the screen. Not all topics have a second level and the user is always informed when a second level is available and is accessible by touching it.

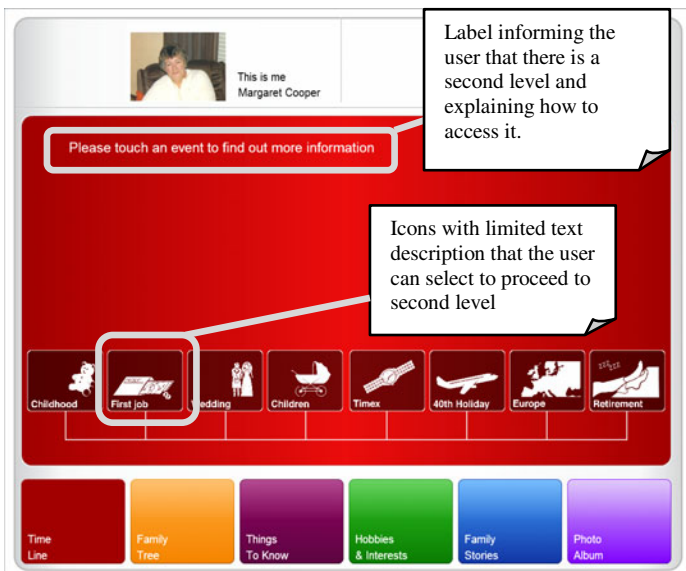


Fig. 1. Time Line topic showing meaningful icons and limited text descriptions that the user can select

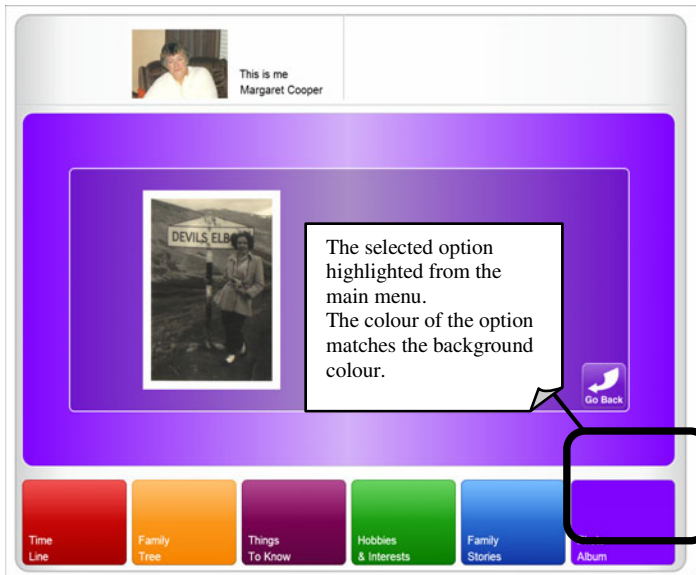


Fig. 2. Selected photo from Photo Album topic with concise text description

3 Evaluation

Method. An initial usability study was carried out to evaluate the system with inexperienced computer users. The study was conducted in two iterations with five participants in each. Iteration 1 had 2 male and 3 female participants. Iteration 2 had 1 male and 4 female participants. Participants varied in age between 18 and 59 years. Most of the participants had never worked in a care environment but two had part-time experience for less than 5 years. Participant's average computer use was daily. Data was collected using a pre and post study questionnaire. The pre-study questionnaire consisted of 5 questions and collected demographic data such as age and computer experience. The post study questionnaire consisted of 17 questions all using a 5-point Likert scale with 1 as the most positive answer and 5 as the most negative answer. The questionnaire was designed to collect data on ease of use, preference of features, level of training required and how engaging Portrait was to use.

During the study audio and screen capture software was used to capture on-screen actions and commentary. This data were used to analyze positive and negative user experiences. As the design of the system was intended to require little training to understand, each participant was only provided with a very brief, introduction to the system (roughly 2 minutes). Participants were then asked to complete three simple training tasks to become familiar with using the system and then to complete five representative tasks that involved finding specific information about the person

profiled in the system (e.g., the names of her children and where she preferred to sit). The participants were asked to ‘think aloud’ while completing the tasks. After completing the tasks the participants were asked to complete the post study questionnaire. Participants were involved in the study for approximately 15 minutes.

Results and Discussion. The first iteration highlighted three usability issues: contrast between background colour and text, font size and the back button being unclear. Two people identified the text as difficult to read due to the lack of contrast between background colour and the text for two topics. Two people said that they would like the font size to be larger so the text would be easier to read. The final issue was elicited from the screen capture where participants were confused by the back button icon (see Figure 3), unsure of its functionality.

Changes to the system to address these issues were made and a second iteration of the study was carried out. The background colour of the Time Line and Things To Know topics was darkened to increase the contrast between the background and the text. The font size was increased from 16pt to 18pt bold for main text and 18pt bold to 20pt bold for headings. The back button icon was changed from the “X” (see Figure 3) to an arrow and label (see Figure 4).



Fig. 3. Original back button with X



Fig. 4. New back button with arrow and label

There were only slight differences (less than 0.5 change in mean) between Iterations 1 and 2 for nine of the 17 questions (see Table 1) so ratings essentially were unchanged and positive. For the remaining questions, Iteration 2 consistently showed more positive ratings than Iteration 1. There were four questions where user ratings changed the most between the two iterations: Level of Training, Size of text, Text font/style and Colour of text (see Table 1 – Q2, Q4.3, Q4.4 & Q4.5). As shown in Figure 5, Level of Training changed from being between 30 minutes and one hour to less than 30 minutes. As seen in Figure 6, participants rated the size, colour and font style of text either very positively or negatively in Iteration 1. In Iteration 2, these attributes were all rated as positive indicating that the changes made to the interface had a positive effect. All three of these questions had the same mean ($M = 2.2$) and standard deviation ($SD = 1.64$) (see Table 1 Q4.3, Q4.4 & Q4.5) in Iteration 1 and Iteration 2 and therefore the same frequency diagram (see Figure 6).

Overall Portrait was very positively received. There were no new usability issues identified as a result of the changes made to the interface for Iteration 2 and all participants rated the system as either engaging or very engaging and fun to use.

Table 1. Mean and standard deviation for post study questionnaire on attitude towards Portrait (rated on a scale of 1 to 5 where 1 is very positive and 5 very negative)

	Usability Study 1		Usability Study 2	
	Mean	SD	Mean	SD
Q1) How difficult it was to learn how to use the system	1.0	0.0	1.0	0.0
Q2) Level of training required to learn how to use the system	2.4	0.55	1.6	0.55
Q3.1) Knowing what to touch for more information	1.4	0.55	1.4	0.55
Q3.2) What information is available in the system	1.6	0.55	1.4	0.45
Q3.3) Reading information on the screen	1.6	0.89	1.2	0.45
Q3.4) Finding the information you wanted	1.8	0.45	1.2	0.45
Q3.5) Changing between topics	1.2	0.45	1.0	0.0
Q3.6) Using the system	1.2	0.45	1.0	0.0
Q4.1) Location of the menu buttons	1.4	0.55	1.0	0.0
Q4.2) Size of buttons	1.4	0.55	1.2	0.45
Q4.3) Size of text	2.2	1.64	1.2	0.45
Q4.4) Text font/style	2.2	1.64	1.2	0.45
Q4.5) Colour of text	2.2	1.64	1.2	0.45
Q4.6) Colour of screen	1.4	0.55	1.2	0.45
Q4.7) Match of the background colour to the button colour	1.4	0.55	1.4	0.55
Q5) Rate how confusing the organisation of information was on the screen	1.4	0.55	1.4	0.55
Q6) How fun was the Profile system to use	1.8	0.45	1.2	0.45

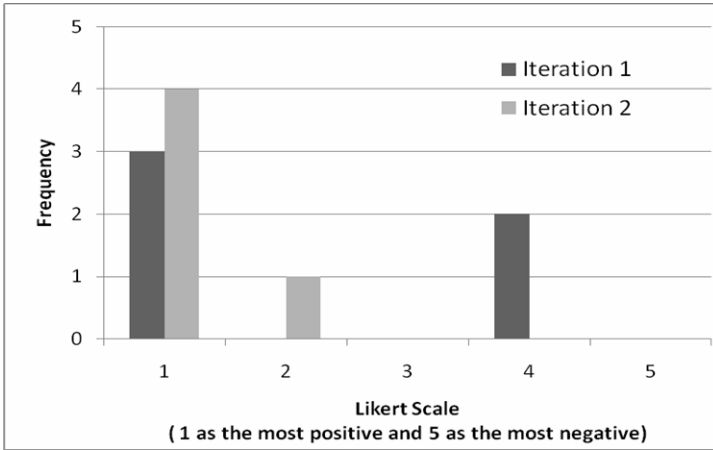


Fig. 5. Frequency diagram of Level of training required to use the system for both iterations of the usability study

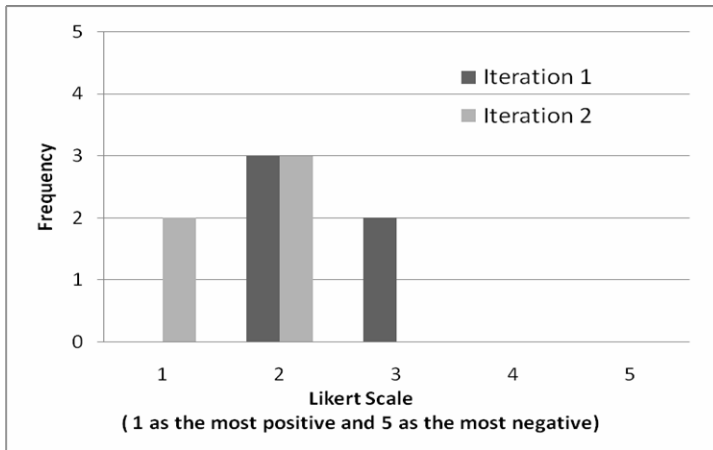


Fig. 6. Frequency diagram of size, colour and font style of text for both iterations of the usability study

4 Conclusion

This paper described a new system called Portrait that is designed to provide personal information and preferences about a person with dementia to care staff. The initial usability study carried out with novice users showed that the Portrait system seemed to be very easy to use and would require little training. The Portrait system had mostly positive outcomes from the first iteration of the usability study with only three usability issues identified. These three issues were corrected for the second iteration and no new usability issues were identified during this second evaluation. These findings suggest that the Portrait could be easily and quickly used by care staff with little computer experience.

The next steps in the research involve populating the Portrait system with the profiles of various fictional individuals with dementia and introducing it to new care staff in actual care homes. We will then investigate the impact that Portrait has on the attitudes of care staff towards the social well-being of patients with dementia. Finally, a series of longitudinal case studies will be carried out with real people with dementia living in care homes to determine the effect of Portrait on the day-to-day social interactions between people with dementia and their care workers over the long term.

Acknowledgments. Funding is provided by the School of Computing, University of Dundee and Balhousie Care Group. We also gratefully acknowledge all of the individuals who participated in the two iterations of the study.

References

1. Ward, R., Vass, A.A., Aggarwal, N., Garfield, C., Cybyk, B.: What is dementia care? Dementia is communication. *Journal of Dementia Care* 13(6), 16–19 (2005)
2. Ward, R., et al.: A different story: exploring patterns of communication in residential dementia care. *Aging & Society* 28, 629–651 (2008)
3. Armstrong-Esther, C.A., Browne, K.D., McAfee, J.G.: Elderly patients: still clean and sitting quietly. *Journal of Advanced Nursing* 19, 264–271 (1994)
4. Caris-Verhallen, W., Kerkstra, A., Bensing, J.: The role of communication in nursing care for elderly people: a review of the literature. *Journal of Advanced Nursing* 25, 915–933 (1997)
5. Kennard, C.: Life Stories. About Inc. (2006), http://alzheimers.about.com/od/professionalresources/a/life_stories.htm (accessed: September 3, 2008)
6. McKeown, J., Clarke, A., Repper, J.: Life story work in health and social care: systematic literature review. *Journal of Advanced Nursing* 55(2), 237–247 (2006)
7. Smith, K.L., Crete-Nishihata, M., et al.: Multimedia Biographies: A Reminiscence and Social Stimulus Tool for Persons with Cognitive Impairment. *Journal of Technology in Human Services* 27(4), 287–306 (2009)
8. Damianakis, T., Crete-Nishihata, M., et al.: The Psychosocial Impacts of Multimedia Biographies on Persons With Cognitive Impairments. *The Gerontologist* 50(1), 23–35 (2009)
9. Alm, N., Ellis, M., Astell, A., Dye, R., Gowans, G., Campbell, J.: A cognitive prosthesis and communication support for people with dementia. *Neuropsychological Rehabilitation* 14(1/2), 117–134 (2004)
10. Alm, N., Dye, R., Gowans, G., Campbell, J., Astell, A., Ellis, M.: A Communication Support System for Older People with Dementia. *Computer* 40(5), 35–41 (2007)
11. Mulvenna, et al.: Reminiscence Systems. In: *HCI 2009*, Cambridge, UK, September 1-5 (2009)
12. Albinsson, P., Zhai, S.: High Precision Touch Screen Interaction. In: *CHI 2003*, April 5-10. Ft. Lauderdale, Florida (2003)
13. Mahmud, M., Kurniawan, S.H.: Involving Psychometric Tests for Input Device Evaluation with Older People. In: *Proceedings of OZCHI 2005*, Canberra, Australia (2005)
14. Sankey, Michael D.: Multiple representations in multimedia materials: an issue of literacy. In: Mishra, Sanjaya and Sharma, Ramesh C., (eds.) *Interactive multimedia in education and training*. IGI Publishing (IGI Global), Hershey, PA, United States (2005)

Mental Models of Menu Structures in Diabetes Assistants

André Calero Valdez¹, Martina Ziefle¹, Firat Alagöz¹, and Andreas Holzinger²

¹ Communication Science, Human Technology Centre, RWTH Aachen University
Theaterplatz 14
52056 Aachen, Germany

² Institute for Medical Informatics, Statistics and Documentation (IMI),
Research Unit HCI4MED, Medical University of Graz
Auenbruggerplatz 2/V
8036 Graz, Austria

{calero-valdez, ziefle, alagoez}@humtec.rwth-aachen.de

Abstract. Demographic change in regard to an aging population with an increasing amount of diabetes patients will put a strain on health care rentability in all modern societies. Electronic living assistants for diabetes patients might help lift the burden on taxpayers, if they are usable for the heterogeneous user group. Research has shown that correct mental models of device menu structures might help users in handling electronic devices. This exploratory study investigates construction and facilitation of spatial mental models for a menu structure of a diabetes living assistant and relates them to performance in usage of a device. Furthermore impact of age, domain knowledge and technical expertise on complexity and quality of the mental model are evaluated. Results indicate that even having a simplified spatial representation of the menu structure increases navigation performance. Interestingly not the overall correctness of the model was important for task success but rather the amount of route knowledge within the model.

Keywords: HCI, Mental Models, eHealth, Diabetes, User Performance, Menu Navigation, Aging, Assisted Living.

1 Introduction

Slick user experience is becoming the critical factor for acceptance, sustainability and competitive capacity of any mobile technical system, especially in regard of demographic changes, world wide increasing life expectancy and the resulting increase of older users. The share of population over the age of 65 already reached 20 percent in Germany in 2008 and is expected to swell to a 38 percent level in 2038[1]. Similar forecasts apply to many western European countries [2]. Usage of electronic devices is also decreasingly voluntary because of either work or everyday life requirements [3,4]. This impact will be even stronger concerning medical appliances of mobile devices. Since the increase of age related illnesses like diabetes accompanies both demographic change and sedentary lifestyle, medical care and age appropriate independent domestic care can only be economically realized through mobile technical solutions [5].

It is a central claim that mobile devices are designed to be in line with users' specificity and diversity. However, the intelligent interface design of mobile devices, which meets the demands and abilities of especially older users, is an extremely sophisticated task. Aging itself represents a highly complex process. Not all users age in the same way, and the onset of aging processes as well as the consequences show considerable differences across humans. Design approaches should therefore take the user-perspective seriously [6-9]. This includes that adults' behavior with current technical devices is carefully studied and also, that user abilities are identified, which affect the interaction with interactive computing devices.

The miniaturization of small screen devices may also contribute to usability shortcomings. Beyond handling and visibility problems, the restricted screen space allows only little information to be displayed at a time. By this, memory load is increased. In addition, orientation in the menu is complicated, because users do not experience how the menu might be "spatially" structured and how the functions are arranged [10-15]. In hierarchically structured menus disorientation occurs when complexity is high with respect to the depth and breadth of menu levels [12, 16, 17].

Earlier research has shown that having a suitable mental model increases usage performance of such devices [11, 18]. The concept of mental models stems from cognitive psychology and is defined as users' internal representation of how systems work and how they are structured [19]. Mental models are assumed to be formed through increasing expertise with the system, combining knowledge of prior experience, cognitive schemata, and problem-solving strategies [20]. Since aging decreases mental processing speed and is accompanied by declines in spatial capability (e.g. mental rotation ability) construction of a spatial mental model could be hampered.

However, it was found that age-related decreases could be compensated by expertise [21, 22]. Thus, performance of older adults can be just as good as that of younger adults when they can rely on elaborated domain-specific knowledge or prior exposure to similar technology.

This study aims to find determining factors of older adults in mental model construction. For this purpose factors such as age, expertise with technology, mobile phone expertise, domain knowledge (in the field of diabetes) were examined as mediating variables on construction of mental models of a diabetes living assistant.

2 Method

In this exploratory study 23 adults were asked to perform five tasks on a simulated diabetes living assistant. Before starting the device, participants were surveyed about their technical expertise in regard to everyday technology (i.e. computers, VCRs, GPS navigation systems, etc.), mobile phones and medical technology. Additionally domain knowledge of diabetes was assessed. After the trials participants were asked to lay out a spatial representation of the diabetes living assistant using screenshots of the program. After this card sorting tasks, participants were asked to shortly comment on their mental models.

Mental models were then grouped into four types (no model = 0, linear = 1, hierarchical = 2, spatial map = 3) regarding their complexity and in regard of their facilitation of spatial ordering cues in the diabetes living assistant. This can be seen as survey

knowledge of the model. Furthermore models were also scored in regard to the expressiveness in route and landmark knowledge [23]. Route knowledge will be rated higher, if the relative distance of two nodes reflects the amount of navigational steps between these nodes. Landmark knowledge is seen as the relative spatial ordering of two screens to each other. A model's accuracy (or quality) is the sum of survey, route and landmark knowledge of the model. Higher score reflects a better incorporation of this knowledge into the model. In order to verify impacts of a constructed model, performance data of the five tasks was related to both type and accuracy of the mental model.

Performance data in this case are effectiveness (percentage of successfully solved tasks) and efficiency. Efficiency measures are the required time for solving tasks, the amount of navigational detour steps and the navigation pace.

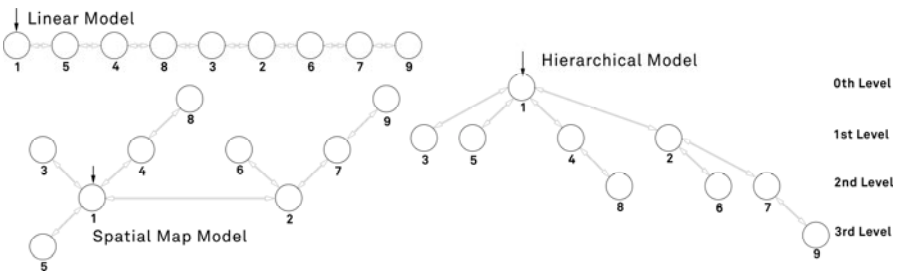


Fig. 1. Linear Model (top left) Hierarchical Model (right) and Spatial Map Model (bottom left)

2.1 Participants

A total of twenty-three adults volunteered to take part in this study. Among those, were seven young adults (M = 27.4 years; SD = 2.6; range: 25 – 33 years), seven medium aged adults (M = 51.3 years; SD = 8.3; range: 41 – 59 years) and nine older adults (M = 67.9 years; SD = 7.8; range: 61 – 87 years). The younger participants were mostly university students of different academic fields. Medium aged and older adults were reached by advertisement in local newspapers and through an exhibition on a local public diabetes convention and covered a broad range of professions and educational levels.

Twelve participants were non-diabetic adults, who were mostly recruited through their social networks. The eleven diabetic participants split up into a group of five participants diagnosed with Type-1-Diabetes and 6 participants suffering from Type-2-Diabetes.

2.2 The Diabetes Living Assistant

Our goal is to develop a portable device that supports diabetes patients in their therapy and in their everyday lives. Before dealing with the hardware part of these devices, we wanted to concentrate on the usability of the software. For the user studies, we needed an application that could be run on standard hardware. Furthermore we needed a user interface that could be used without a keyboard, so that it can later be

used on a touch screen-enabled mobile device. Another constraint was given by our decision to use the Jacareto capture and replay toolkit [22, 24, 25].

Instead of creating a specialized application that is only useful for certain diabetes patients, we decided to include features that are required for the different types of the disease.

We have tried to keep the menu-depth and breadth at a minimal level by using a combination of a hierarchical menu and a wizard-like linear structure for data input (see Fig. 2 – node 3). The main menu consists of two screens (nodes 1 and 2) with 4 (or 3 respectively) function keys. In order to improve ease of spatial mapping each function key is placed in one corner of the screen, to suggest a spatial placement of the feature behind the function.

The diabetes living assistant was simulated as a software solution on a PC connected to a touch screen, which was masked with an opaque paper cutout to create a device like feel.

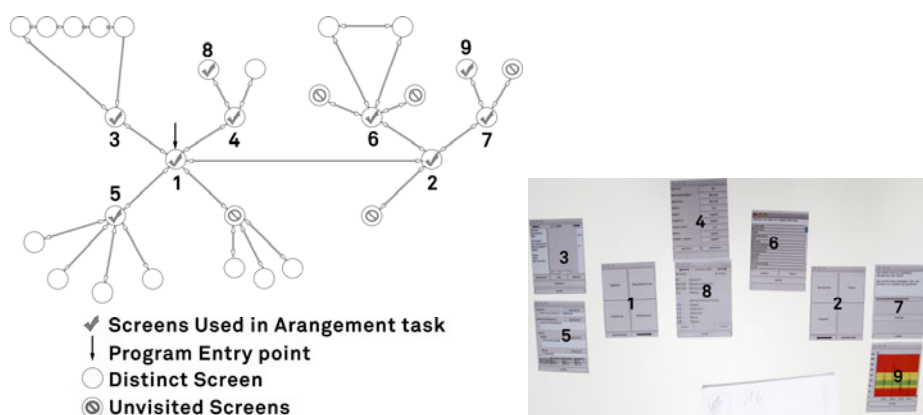


Fig. 2. Simplified Navigational Map of the Diabetes Living Assistant (left), Layout of Printed Screenshots (with a number overlay) as a Spatial Map (right)

3 Results

Results of this study were analyzed by one-way ANOVA, bivariate correlations, multivariate analysis of variance and univariate analysis of covariance with a level of significance set at 5%.

3.1 Did Participants Elicit the Model of the Programm Correctly?

Participants showed four different general types of mental maps. Only three older users were completely unable to arrange any model at all. Four users ordered the screens into a linear model, six into a hierarchical and eleven users arranged a spatial map as a mental model (see Table 1).

Table 1. Distribution of Models in Regard to Age Group, ordered from left to right in increasing complexity

	No Model	Linear Model	Hierarchy Model	Map Model
Age 25-33	0	0	2	5
Age 41-59	0	2	0	5
Age 60-87	3	1	4	1

Correlation analyses show that age is highly significantly correlated with the complexity of the selected mental model ($r = -0.566$) as well as with the quality of the arranged model of the user ($r = -0.467$). The quality of the arranged model is also significantly correlated to the users experience with mobile phones ($r = -0.430$). Interestingly no significant correlation was found between complexity and quality of model with other technical expertise or domain knowledge.

3.2 How Did Personal Factors Impact Performance?

Age and expertise with technology show a significant correlation with performance measures in analyses. Younger age is highly correlated with better effectiveness ($r = -0.664$) and efficiency. Younger users need less total steps ($r = 0.616$), make less navigation errors ($r = 0.472$) and have a faster navigation pace ($r = 0.693$). Expertise with technology is mostly correlated with effectiveness ($r = -0.449$) such that users with better expertise are more effective than users that are more inexperienced. This correlation does only affect one efficiency measurement significantly (i.e. total time $r = 0.476$), which also shows that higher expertise is related with better performance.

3.3 Did Participants with More Accurate Models Perform Better?

Correlation analyses show that, having a general correct assumption of the complexity of the underlying spatial model of the program, is significantly correlated with task effectiveness ($r = 0.496$) and the navigational pace ($r = -0.453$, lower value is faster speed) in the menu of the device. The accuracy of the model only significantly correlates to the overall task effectiveness ($r = 0.473$).

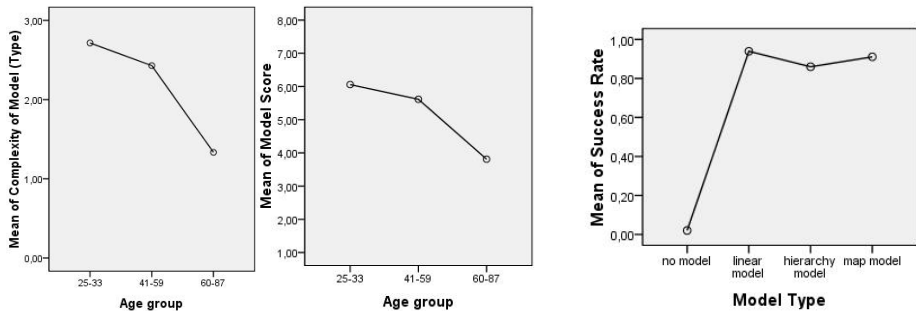


Fig. 3. Comparison of Mental Models in Regard to Complexity (left) and Quality (middle), Mean of Success Rate over Type of Model (right)

Still, one-way ANOVA comparison of means shows that not the type of the model alone can explain performance differences since users with a linear model performed equally well as users with a hierarchical or spatial model (see Figure 3). Linear Regression shows that route knowledge has the biggest impact on performance (22%) and outweighs survey knowledge (16%) and landmark knowledge (7%). Users who had a linear mental model incorporated about the same amount of route knowledge into the models as the users with hierarchical models (mean difference 0.166).

4 Discussion and Conclusion

Not only do several factors affect performance of users during handling of a small screen device, but the mental model and especially lack thereof can heavily impact whether initial usage of a device is successful or has a propensity for failure. Age and lack of experience with mobile phones decrease the ability to create a correct or useful mental model. Interestingly not the overall correctness of the model was important for task success but rather the amount of route knowledge within the model. Users who arranged a linear model performed equally well as users with hierarchical and spatial map models (see Figure 3). Two possible explanations directly come to mind. Either the users arranged the screens by mapping their model (e.g. by traversing a hierarchy) onto a linear one or they truly created a mental map in a linear fashion.

Two participants with a linear mental model reported during the experiment that there were insufficient printouts of the main menu screens, so they tried to arrange them as good as possible, knowing there were missing connections between some screens and the menu screens. This leads to the assumption that they were not reusing the menu screens (like in a spatial hierarchy) and rather saw them as individual instances of a menu on a timeline based mental model.

Surprisingly, domain knowledge did not significantly impact the quality of the mental model, and as a consequence, the navigation performance. Thus, diabetic users could not profit from their more profound medical knowledge about what to do within a diabetes illness, when navigating through the device menu. In contrast, the expertise with mobile phone showed to have an enormous importance for the formation of a useful mental model.

4.1 Applications of This Study

The results of this study have shown that not all users might require a perfect mental model in order to operate a diabetes living assistant. Still, having a model at all is crucial to succeeding, corroborating findings of earlier studies [4][13]. Older users that are not familiar with hierarchical menu structures might be unable to elicit such a model from using a program. Since this study was exploratory in nature, results need to be validated with research. As the findings refer to a relatively small group of participants, we have to be cautious to transfer the outcomes to the broad variety of older users. Additionally the quality of specific parts of the mental model should be analyzed in regard to performance of tasks that require only aforementioned parts of the program. Designing a menu structure in a pure linear fashion, or in a way that it can easily be traversed into a linear mental model, could make usage easier for elderly

patients. Showing a user a spatial overview of the program might also yield increased performance [11][13], due to priming of a mental model.

Acknowledgments. This work was funded by RWTH Aachen postgraduate scholarships for students with excellent degrees and by the excellence initiative of the German federal and state governments.

References

1. Siebert, H.: *Economic Policy for Aging Societies*. Springer, Berlin (2002)
2. Federal Statistical Office Germany: *Germany's population until 2050 – 11th coordinated population projection*, Wiesbaden (2006), <http://www.destatis.de/>
3. Arning, K., Ziefle, M.: Understanding age differences in PDA acceptance and performance. *Computers in Human Behaviour* 23, 2904–2927 (2007)
4. Arning, K., Ziefle, M.: Barriers of information access in Small-Screen-Device applications: The relevance of user characteristics for a transgenerational design. In: Stephanidis, C., Pieper, M. (eds.) *ERCIM Ws UI4ALL 2006*. LNCS, vol. 4397, pp. 117–136. Springer, Heidelberg (2007)
5. Leonhardt, S.: Personal Healthcare Devices. In: Mukherjee, S., et al. (eds.) *AmIWare: Hardware Technology Drivers of Ambient Intelligence*, pp. 349–370. Springer, Dordrecht (2006)
6. Holzinger, A.: Usability Engineering for Software Developers. *Communications of the ACM* 48(1), 71–74 (2005)
7. Holzinger, A., Searle, G., Kleinberger, T., Seffah, A., Javahery, H.: Investigating Usability Metrics for the Design and Development of Applications for the Elderly. In: Miesenberger, K., Klaus, J., Zagler, W., Karshmer, A. (eds.) *ICCHP 2008*. LNCS, vol. 5105, pp. 98–105. Springer, Heidelberg (2008)
8. Nischelwitzer, A., Pintoffl, K., Loss, C., Holzinger, A.: Design and Development of a Mobile Medical Application for the Management of Chronic Diseases: Methods of improved Data Input for Older People. In: Holzinger, A. (ed.) *USAB 2007*. LNCS, vol. 4799, pp. 119–132. Springer, Heidelberg (2007)
9. Holzinger, A., Searle, G., Nischelwitzer, A.K.: On some Aspects of Improving Mobile Applications for the Elderly. In: Stephanidis, C. (ed.) *HCI 2007*. LNCS, vol. 4554, pp. 923–932. Springer, Heidelberg (2007)
10. Ziefle, M., Bay, S.: How older adults meet complexity: aging effects on the usability of different mobile phones. *Behaviour & Information Technology* 24(5), 375–389 (2005)
11. Ziefle, M., Bay, S.: How To Overcome Disorientation in Mobile Phone Menus: A Comparison of Two Different Types of Navigation Aids. *Human-Computer Interaction* 21, 393–433 (2006)
12. Lin, D.M.: Age differences in the performance of hypertext perusal. In: *Proceedings of the Human Factors and Ergonomics Society 45th Annual Meeting*, pp. 211–215. Human Factors Society, Santa Monica (2001)
13. Ziefle, M., Bay, S.: Mental Models of a Cellular Phone Menu. Comparing Older and Younger Novice. In: Brewster, S., Dunlop, M. (eds.) *Category Theory Applied to Computation and Control*. LNCS, pp. 25–37. Springer, Heidelberg (1975)
14. Ziefle, M.: Aging and Mobile Displays: Challenges and requirements for age-sensitive electronic information designs. In: *Proceedings of the 9th International Conference on Work With Computer Systems, WWCS 2009, Beijing, China (2009)*

15. Arning, K., Ziefle, M.: Assessing computer experience in older adults: Development and validation of a computer expertise questionnaire for older adults. *Behaviour and Information Technology* 27(1), 89–93 (2008)
16. Parush, A., Yuviler-Gavish, T.: Web navigation structures in cellular phones. The depth-breadth trade-off issue. *International Journal of Human-Computer Studies* 60, 753–770 (2004)
17. Westermann, S.J.: Individual differences in the use of command line and menu computer interfaces. *International Journal of Human Computer Interaction* 9, 183–198 (1995)
18. Holzinger, A., Schaupp, K., Eder-Halbedl, W.: An Investigation on Acceptance of Ubiquitous Devices for the Elderly in an Geriatric Hospital Environment: using the Example of Person Tracking. In: Miesenberger, K., Klaus, J., Zagler, W., Karshmer, A. (eds.) *ICCHP 2008*. LNCS, vol. 5105, pp. 22–29. Springer, Heidelberg (2008)
19. Gentner, D., Stevens, A.L.: *Mental models*. L. Erlbaum Associates, Hillsdale (1983) (Gentner, D., Stevens, A.L. (eds.))
20. Sasse, A.: Cognitive Ergonomics and Human-Computer Interaction. *International Journal of Man-Machine Studies* 35(2), 269–270 (1991)
21. Morrow, D., Miller, L.S., Ridolfo, H., Kokayeff, N., Chang, D., Fischer, U., Stine-Morrow, E.: Expertise and aging in a pilot decision making task. In: *Proceedings of the Human Factors and Ergonomics Society 48th Annual meeting*, pp. 228–232 (2004)
22. Calero-Valdez, A., Ziefle, M., Horstmann, A., Herding, D., Schroeder, U.: Effects of aging and domain knowledge on usability in small screen devices for diabetes patients. In: Holzinger, A. (ed.) *USAB 2009*. LNCS, vol. 5889, pp. 366–386. Springer, Heidelberg (2009)
23. Thorndyke, Goldin: Spatial learning and reasoning skill. In: Pick, H.L., Acredolo, L.P. (eds.) *Spatial orientation. Theory, research, and application*, pp. 195–217. Plenum, New York (1983)
24. Schroeder, U., Spannagel, C.: The Role of Interaction Records in Active Learning Processes. In: Isaias, P., Nunes, M.B., dos Reis, A.P. (eds.) *Proceedings of the IADIS Virtual Multi Conference on Computational Information Systems*, pp. 99–104; *Behaviour* 21, 195–209 (2005)
25. Spannagel, C., Gläser-Zikuda, M., Schroeder, U.: Application of Qualitative Content Analysis in User-Program Interaction Research. *Forum: Qualitative Social Research*, 6(2) (2006), <http://www.qualitative-research.net/fqs-texte/2-05/05-2-29-e.htm>

Touch Screen User Interfaces for Older Subjects

Effect of the Targets Number and the Two Hands Use

Guillaume Lopicard and Nadine Vigouroux

Université Paul Sabatier, Institut de Recherche en Informatique de Toulouse,
118 route de Narbonne, 31062 Toulouse Cedex 9, France
{lepicard,vigourou}@irit.fr

Abstract. This study investigated the optimal numbers of blocks and tactile targets for touch screen user interfaces intended for use by all (reference and older population). Three independent variables (number of targets, number of interaction blocks on the touch screen and number of hands used) were studied in our experiment. Huge amount of data were stored. In this paper, we will only report statistical analyses on two variables: Time needed to Realize the Test and Error Rate. Each variable will be analyzed in two times: for the whole population (reference and older) and for the comparison between these two populations.

Keywords: older people, touch screen, interface design, bi-manual.

1 Introduction

In 2020, 20.06% of the population in the European Union will be aged 65 years or over. The number will grow to 29.95% in 2060 [1]. The motivation is in the demographic change and ageing in Europe, which implies not only challenges but also opportunities for the citizens, the social and healthcare systems through the use of Information and Communication Technologies (ICT). The computer human interaction must be well designed and must satisfy usability criteria for elderly people. Some researches based on the use of ICT reported that the computers based systems [3] (computers, smart phones, electrical domestic integrated touch screen, etc.) will be more and more available in the house of older adults. Charparro and al. [5] and Wood and al. [10] have shown that the handling of computer with a keyboard and mouse can cause problems with older people. Touch screen are widely used in home systems for environment control and health care. The new generation of touch screen offers multi touch interaction. This interaction was democratized by the iPhone of Apple.

The intuitiveness and the strong affordance in the use of fingers without ICT could be acceptable in interface for older people. Woods and al. [10] reported that touch screen offers more intuitive modes of interaction (direct input) and requires less visual spatial coordination. Jin and al. [6] gave some guidelines for button size and spacing for older subjects. Numerous studies were realized on two hands interaction [4], [7]. But, the use of two hands for older people was not enough studied.

We will describe in this article the experiment and will present our results. These results will permit us to determine the variables that impact the user performance whatever the population. We analyze the time of realization and the error rates in function of number of hands, number of blocks and number of targets for two populations (reference people and older people).

2 Experiment Design

2.1 Subjects

Thirty-six people aged from 15 to 52 years (mean= 32.2, SD= 10.1) participated as subjects in a first experience. Twenty-four older adults aged from 63 to 88 years (mean= 76.5, SD= 8.2), without cognitive impairment, were recruited for the experiment. For all subjects, the visually impairment was corrected by wearing glasses if necessary. We compared the two populations to identify the impact of parameters on age.

2.2 Interface Design

The interface (Figure 1a, 1b) is divided in three parts. The first part shows the number of user and the hands number that the user has to realize the test. It situates in the upper left corner of screen. The second is located at the center of the screen; it is the round target that indicates the color that should be recognized. The last part is the area of the targets at the bottom of the screen.

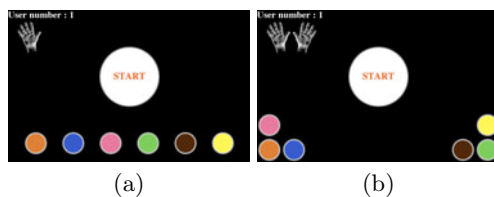


Fig. 1. (a) Test with 1 hand on the layout: 6 targets and 1 block, (b) Test with 2 hands on the layout: 6 targets and 2 blocks

2.3 Targets

Target size measures 100 x 100 pixels corresponding to 2 cm x 2 cm on screen. This is the target size recommended to have faster response performance by older subjects [6] on touch screen. Targets are colored with one of height named colors (red, green, blue, orange, yellow, brown, pink, purple) [9] and have a round shape. They are surrounded by a gray gradient outline of 10 pixels. The instruction circle measures 300 x 300 pixels and 6 cm x 6 cm on screen. The target disposition consists of one or two interaction blocks (Figure 1a, Figure 1b). Using one block, the targets are in line with a uniform distribution and

situated at the bottom of screen. Using two blocks, the targets are divided in two parts situated at the bottom of screen, one on the left and the other on the right. The color/position pair does not change for a same layout (number of targets and number of blocks fixed).

2.4 Task

The subjects were invited to realize a trial practice. Firstly, we had assured that the subject understood the goal statement and that the identification (perception and recognition) of colors was correct. For the test, the task of user was to touch the color (target) corresponding to the instruction, as quickly and accurately as possible, with one finger by hand. The instruction changed each five seconds if the user did not click on the right target. If the user clicked on wrong target, nothing happened.

2.5 Procedure

Each participant realized twelve tests on twelve different configurations (one or two hands on each of the six layouts: one or two blocks / 4, 6 or 8 targets). We had to have a pair number of targets and less than 9 targets [8]. The case of two targets was not considered because of its ease. The tests were divided in two balanced groups according to the number of hands (6 tests with one hand then 6 with two). To minimize learning and ordering effects, the order of the tests (number of blocks and number of targets) were balanced. For each test, the user clicked 48 times (4, 6 and 8 targets respectively 12, 8 and 6 clicks).

Following the targets number, the order of apparition was randomized between 50 different ways. All the ways (4, 6 and 8 targets) were equivalent on the 1D distance. The same color did not consecutively appear. Each color had the same occurrence number on the instruction circle. The number of clicked targets was equal to 576 (12×48). To start the tests, the participant was invited to touch on the start button. The user study took about 20 to 25 minutes per reference participant and 25 to 60 minutes per older participant.

2.6 Equipments

Participants used a Dell Tablet PC Latitude XT2, with a 12.1" inch LCD monitor running Microsoft Windows 7 operating system. The screen size was 1280 x 800 pixels resolution (124,7dpi). The virtual keyboard was disabled. The Tablet was defined with the tactile interaction. It was placed on a desk with an inclination of 30 degrees [2]. The subjects were invited to sit in front of the desk and to adjust their visual position (to be at a comfortable viewing distance and avoid reflection of the touch screen). The application was implemented on Adobe Flash and use amfphp to communicate with the mysql database.

3 Hypotheses

We have formulated four hypotheses divided on two parts.

Two global hypotheses: Hypothesis 1: The two subject populations (reference and older) would be slower with one hand and two blocks than the other combinations while with two hands they would be faster; Hypothesis 2: The error rate would be equivalent whatever the configuration.

Two hypotheses related to the difference between reference and older people: Hypothesis 3: The older people would be slower than the reference people to realize the test; Hypothesis 4: The older people would realize more errors than reference people.

4 Results

We have defined two parameters: the Time needed to Realize the Test (TRT) and the Error Rate (ER). The TRT is the total time to realize completely the test within a specific configuration. The ER is the error rate for each configuration, it may be upper than hundred per cent: the number of errors divided by the number of clicks to realize the test. If the subject makes two errors (whatever the type of errors) for each instruction, the error rate will be equal to two hundred per cent.

The results were analyzed in two steps: firstly, the mix of all people; secondly, the comparison of population between reference and older population. For each variable analyzed, a Shapiro test and a Bartlett test were realized. If the homoscedasticity condition was respected, an ANOVA was realized. Otherwise, a Kruskal-Wallis test, and if the test was significant, a Nemenyi-Damico-Wolfe-Dunn (NDWD) test was realized to compare the different means.

4.1 The Time Needed to Realize the Test

Global results. Three way ANOVA analysis of TRT revealed that the mean of TRT significantly depended on the number of targets ($F(2,708)=21.4943$, $p=8.66e-10$). A trend existed for the interaction with the number of hands and the number of blocks ($F(1,708)=2.7677$, $p=0.09663$). When we fixed the number of hands and the number of blocks, significant difference between four targets and eight targets ($p=0.002772344$) for the combination of 1 hand / 1 block ($\text{Chi}2(2)=10.9839$, $p=0.00412$). For the combination of 1 hand / 2 blocks, there is significant difference appeared ($F(2,177)=7.491$, $p=0.0007535$). The TukeyHSD test indicated a difference between four and eight targets ($p=0.0004503$). For the combination of 2 hands / 1 block, significant difference existed ($\text{Chi}2(2)=13.9711$, $p=0.0009252$). The NDWD test showed a difference between four and eight targets ($p=0.0005588741$). Finally, for the combination of 2 hands / 2 blocks; significant difference appeared ($F(2,177)=6.5339$, $p=0.001829$). The TukeyHSD test indicated difference between four and eight targets ($p=0.0012394$). The interaction between the number of hands and the number of blocks (Figure 2a) explicated by the fact that when the subjects have only one hand to realize the configuration with two blocks, the distance is longer than the other configurations.

Comparison between population. Four way ANOVA analysis of TRT showed significant differences for: the number of targets ($F(2,696)=52.8997, p < 2.2e - 16$), the number of blocks ($F(1,696)=6.1545, p=0.0133433$) and the type of population ($F(1,696)=1017.6496, p < 2.2e - 16$). Few interactions existed between parameters: number of hands with number of blocks ($F(1,696)=6.8115, p=0.0092518$); number of hands with type of population ($F(1,696)=5.3634, p=0.0208539$); and number of targets with type of population ($F(2,696)=9.0447, p=0.0001325$). For the analysis of global results interaction (Figure 2a), the 1 hand / 2 blocks configuration had a TRT longer than other configurations. The interaction between the number of hands and the type of population (Figure 2b) indicated that the use of two hands by older people is more difficult than by reference people. The interaction between the number of targets and the type of population (Figure 3) showed that the complexity of configuration (here the number of targets) had a greater effect for the elderly than the reference people.

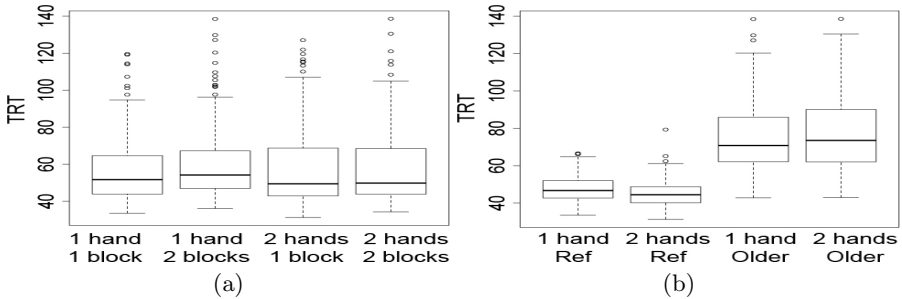


Fig. 2. (a) Impact of hands number and blocks number on TRT, (b) Impact of hands number and type of population on TRT

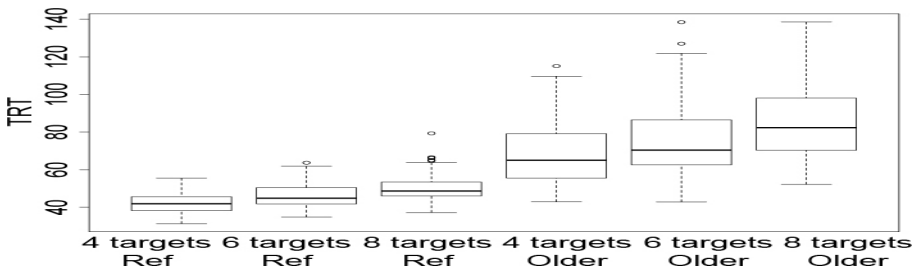


Fig. 3. Impact of targets number and type of population on TRT

4.2 The Error Rate

Global results. Three way ANOVA analysis of ER revealed that the mean of ER was significantly different for the number of targets ($F(2,708)=3.1702,$

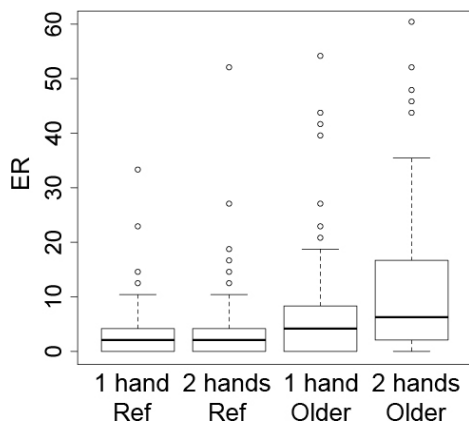


Fig. 4. Impact of hands number and type of population on ER

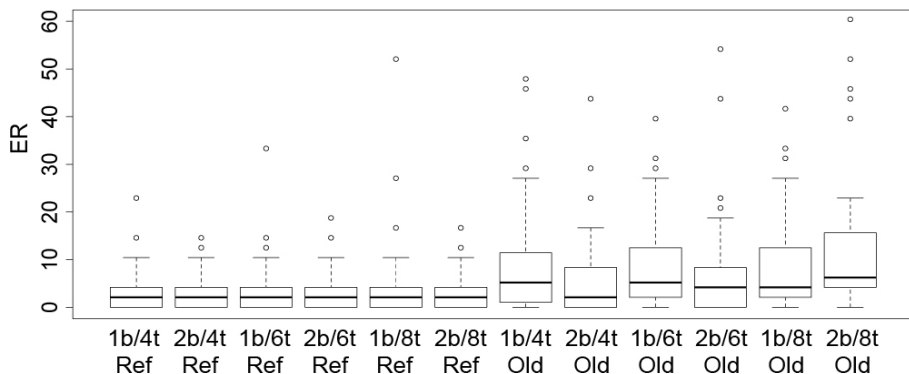


Fig. 5. Impact of blocks number, targets number and type of population on ER (ex: 1b/4t Ref = 1 block / 4 targets / Reference people)

$p=0.04259$) and the number of hands ($F(1,708)=16.6631, p=4.973e-05$). When we fixed the number of hands and the number of blocks, we only discovered a trend between four and eight targets ($p=0.0660038$) for the combination 2 hands / 2 blocks ($F(2,177)=2.9445, p=0.05521$). When we fixed the number of targets and the number of blocks, we detected significant differences within configurations: 1 block / 4 targets ($F(1,118)=5.39, p=0.02197$), 1 block / 6 targets ($F(1,118)=4.0024, p=0.04773$) and 1 block / 8 targets ($F(1,118)=4.7392, p=0.03147$).

Comparison between population. Four way ANOVA analysis of ER showed that differences appeared for: the number of hands ($F(1,696)=19.4068, p=1.223e-05$); the number of targets ($F(1,696)=3.6922, p=0.0254064$); and the type of population ($F(1,696)=94.8222, p < 2.2e - 16$). Few interactions existed: between the

number of hands and the type of population ($F(1,696)=13.7078$, $p=0.0002305$); and between the number of blocks, the number of targets and the type of population ($F(2,696)=4.5623$, $p=0.0107525$). The interaction between the number of hands and the type of population (Figure 4) indicated that the use of two hands for older people had strongly impact on the error rate contrary to reference people. The interaction between the number of blocks, the number of targets and the type of population (Figure 5) showed that for older people, the number of blocks had impact on the ER contrary to reference people. However, the impact was different according to the number of targets: for 4 and 6 targets, the ER was a little lower with two blocks than with one block, inversely for 8 targets. This can be explained by the colors and their arrangement toward other colors.

5 Discussions

Statistical analysis gave a lot of information on the performance of subject. These also permit us to respond to some hypotheses.

H1: This hypothesis was wrong. The three way ANOVA showed that the only parameters which had significantly impact on the TRT was the number of targets. The configurations with eight targets were longer than that with four targets. Generally, more the number of targets increased, more the TRT rose. On the other side, the hypothesis on the configuration 1 hand / 2 blocks was right, this configuration was longer than the others.

H3: This hypothesis was true. For all configurations the difference between the two populations was true. We can conclude that the age influences the TRT. We can observe that, contrary to reference people, the use of two hands for older people increases the TRT. So, the number of targets had consumed more cognitive resources by older people than by reference people.

H2: This hypothesis was wrong. The number of hands and the number of targets showed significant differences on the error rate. In the configuration with one block, having two hands increases the error rate whatever the number of targets. We can deduce that it is not advised to use two hands with an interface built in one block.

H4: This hypothesis was true. Generally, older people make more errors than reference people. It was particularly true for the configuration with two hands, the ER increased strongly for older people by comparison with reference people. The number of blocks had impact on ER with older people. But, this impact was different following the number of targets. This difference can be explained by the position of colors.

6 Conclusion and Planned Activities

Our study permits to show that the use of two hands lead to increase the TRT and ER for older people and consequently that it is not recommended to design an interface with the use of two hands for older people. This study shows that it is important to reduce the cognitive overload in the interface if we want great

performance for older people. It advises to limit the number of targets and the number of blocks.

Future works are planned to increase the size of two populations, to analyze the results according little age intervals (by step of five) and to model the effect of age on the performance (speed, error rate, etc.). Then, we hope that tactile recommendations could be given for the design of touch screen interface.

References

1. Eurostat. Technical report, http://epp.eurostat.ec.europa.eu/cache/ITY_OFFPUB/KS-SF-08-072/EN/KS-SF-08-072-EN.PDF
2. Ahlström, B., Lenman, S., Marmolin, T.: Overcoming touchscreen user fatigue by workplace design. In: Posters and short talks of the 1992 SIGCHI conference on Human factors in computing systems, Monterey, California, pp. 101–102. ACM, New York (1992)
3. Beach, S., Schulz, R., Downs, J., Matthews, J., Barron, B., Seelman, K.: Disability, age, and informational privacy attitudes in quality of life technology applications: Results from a national web survey. *ACM Trans. Access. Comput.* 2(1), 1–21 (2009)
4. Buxton, W., Myers, B.: A study in two-handed input. *SIGCHI Bull.* 17(4), 321–326 (1986)
5. Chaparro, A., Bohan, M., Fernandez, J., Choi, S.D., Kattel, B.: The impact of age on computer input device use: Psychophysical and physiological measures. *International Journal of Industrial Ergonomics* 24(5), 503–513 (1999)
6. Jin, Z., Plocher, T., Kiff, L.: Touch screen user interfaces for older adults: Button size and spacing. In: Stephanidis, C. (ed.) *HCI 2007. LNCS*, vol. 4554, pp. 933–941. Springer, Heidelberg (2007)
7. Latulipe, C., Mann, S., Kaplan, C.S., Clarke, C.L.A.: symSpline: symmetric two-handed spline manipulation. In: Proceedings of the SIGCHI conference on Human Factors in computing systems, Montréal, Québec, Canada, pp. 349–358. ACM, New York (2006)
8. Miller, G.: The magical number seven, plus or minus two: Some limits on our capacity for processing information (1956), <http://cogprints.org/730/>
9. Vandenbroek, E., Schouten, T., Kisters, P.: Modeling human color categorization. *Pattern Recognition Letters* 29(8), 1136–1144 (2008)
10. Wood, E., Willoughby, T., Rushing, A., Bechtel, L., Gilbert, J.: Use of computer input devices by older adults. *Journal of Applied Gerontology* 24(5), 419–438 (2005)

Using a Wearable Insole Gait Analyzing System for Automated Mobility Assessment for Older People

Johannes Oberzaucher¹, Harald Jagos¹, Christian Zödl¹,
Walter Hlausehek¹, and Wolfgang Zagler^{1,2}

¹ CEIT RALTEC, Am Concorde Park 2, 2320 Schwechat, Austria

{j.oberzaucher,h.jagos,c.zoedl,w.hlausehek,w.zagler}@ceit.at

² Fortec – Research Group on Rehabilitation Technology, Vienna Univ. of Technology,
Favoritenstrasse 11/029, A-1040 Vienna, Austria
zagler@fortec.tuwien.ac.at

Abstract. Falls among the older population are one of the most common causes for injuries, frailty and for morbidity. Fall incidents have various reasons and are often related to decreased mobility and hence an increasing fall risk could be detected in time. The objective of this paper is to show results and future prospects of the funded project “*vitaliSHOE*” and in detail of an *automated multi-sensor-based method* to determine fall risk indicators in older people's gait and body movements.

Keywords: wearable, fall risk assessment, gait analysis, accelerometer, ambient assisted living.

1 Introduction

Several clinical and mobility assessment tools for fall risk estimation have been developed and evaluated in the last decade; a good summary can be found in [1]. However, most of these tests have to be performed in the presence of an observer and mostly in clinical settings. Thus they suffer from the problem of the inter-observer variability. Hence several studies have been performed to compensate this problem by following the idea of an extramural mobility assessment [2,3,4]. Related projects have shown the ability of various sensor-systems to extract significant parameters for detecting an increased fall risk during performing mobility assessment tests [2].

In our approach a self developed insole gait measurement system is used to determine spatio-temporal gait parameters during specified assessment procedures. This embedded insole system offers the opportunity to measure the bipedal gait and body movements wirelessly and directly at the foot in almost any surrounding while wearing normal walking shoes. By combining and fusing different sensor-information a multi-modal presentation of gait and foot movements can be assessed. From a set of specific movements significant parameters can be extracted in a subsequent automated gait analysis procedure.

2 Methods

2.1 User Involvement

As preliminary work for ongoing R&D work extensive expert and user involvement was carried out. On the one hand an evaluation of possible and significant fall risk assessment methods had to be done. On this account interviews with medical experts and physiotherapist as well as with gait laboratories were carried out to form the basis for further R&D work. On the other hand a pool of older test-participants had to be created, always following the principles of a continuous user involvement throughout the whole project.

2.2 Mobility Assessment and Gait Parameter Measurement

For our approach common clinical mobility assessment tools are used, two of them will be specified in this paper. These assessment tests, namely the Timed “Up and Go” (TUG) and the “Sit To Stand 5” (STS5) test offer a high significance in estimating older people's risk of falling [1].

While test participants perform the specified assessment movements, sensor data from both feet are recorded at sampling rate of 200 Hz and are internally pre-processed in both embedded insole systems. The processed data sets are then transmitted wirelessly to a base station and the spatio-temporal parameters are recorded with 50 Hz by a software running on a Windows system.

Afterwards the recorded gait data are synchronized, cleaned and analysed using a toolbox consisting of scripts written by the authors in Matlab. In a next step an algorithm developed during the project is used for automated extraction of significant assessment parameters.

3 R&D Work and Results

3.1 Wearable Insole Measurement System

For extramural gait analysis and following the vision of a self-administrable mobility assessment a gait measurement system was developed, which offers the opportunity of measuring bipedal spatio-temporal gait parameters with a sampling rate of 50 Hz (e.g. gait times, parameters from a 6-DOF inertial measurement unit, plantar pressure distributions) (Fig. 1). The multi sensory measurement system was validated during a laboratory evaluation in a gait laboratory at the University for Applied Sciences in Vienna (FH Technikum Wien) in 2009 and 2010.

3.2 Gait Assessment and Automated Feature Extraction

During the first project phase an algorithm for automated extraction of significant gait and mobility parameters based on data recorded in laboratory settings was developed. After this first phase initial user involvement tasks and gait monitoring with older

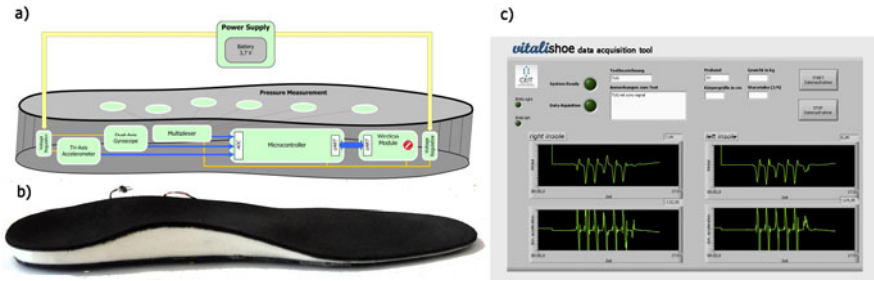


Fig. 1. a) Concept of the insole system, b) Final insole with integrated hardware, c) Gait data acquisition software

people were performed in the Living Lab Schwechat at the Centre for Elderly Citizens in Schwechat in November 2009. Data from 3 male subjects (n=3, age>60; two with normal fall risk, one with increased fall risk) was collected and used to verify the developed feature extraction algorithms.

By using extensive signal processing and the gait-signal analysis algorithm single features (n=9) from the different assessment tests as well as spatio-temporal features (n=24) from the older people's gait have been *automatically* extracted. An Example for the automated assessment feature extraction of TUG parameters is shown in Fig. 2. So, in addition to the common temporal parameters for the mobility tests (n=2) a broad amount of new parameters can be extracted, which are likely to correlate with decreasing mobility and increased fall risk [3].

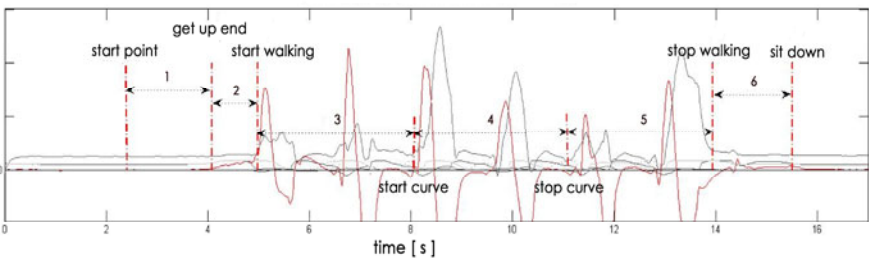


Fig. 2. Multi-sensor presentation of a TUG - Example of Automated Assessment Feature Extraction in a Timed "Up and Go" Test. Extracted Parameters: Full Time, Get Up Time (1), Start Walk Time (2), Straight Walk Time(3 and 5), Curve Time(4), Sit Down Time(6).

4 Impact or Contribution to the Field

Falls and fall-related injuries in the older population are one of the major causes for frailty and due to that for morbidity. They are one of the main reasons for increasing costs in the health care system in developed countries. Up to now many studies have investigated fall risks in controlled laboratory settings, but only a few have investigated these fall risks in real-life settings.

By using our extramural gait analysis system for fall risk assessment a broad amount of new features characterizing older people's abilities can be assessed. Based on these results, classification methods can be used for better and faster prediction of fall risk. These results mark a contribution towards a better and earlier fall prevention and by this a chance for a longer self-determined life at home.

5 Conclusion and Planned Activities

A small and unobtrusive gait measurement system was developed, serving the purpose of monitoring gait and body movements during performing mobility assessments in an unsupervised setting in any surrounding. The system can be used to record gait and body movement data in order to extract significant parameters for fall risk estimation.

In a next step a small clinical trial with a collective of fallers and non-fallers will be conducted, in order to show the significance of the proposed system for classification tasks in the area of fall risk assessment and to evaluate the results against an external reference system.

The results and developed algorithms will be integrated into a JAVA framework and implemented in the vitaliSHOE system in order to realize an unsupervised and usable fall risk assessment for older people in their own homes. The project received funding from the Austrian Research Promotion Agency FFG (program line *benefit*).

References

1. Oliver, D., et al.: Risk factors and risk assessment tools for falls in hospital in-patients: a systematic review. *Age and Ageing* 33, 122–130 (2004)
2. Narayanan, M., et al.: Fall Management: Detection and Prevention, using a Waist-mounted Triaxial Accelerometer. In: 29th Annual International Conference of the IEEE EMBS, Lyon, France (2007)
3. Lord, S.R., et al.: Streamlining Assessment and intervention in a falls clinic using Timed Up and Go Test and Physiological Profile Assessment. *Age and Ageing*, 567–571 (2005)
4. Mathie, M.J., et al.: A pilote study of long-term monitoring of human movements in the home using accelerometry. *J. Telemed Telecare* 2004 10, 144–151 (2004)
5. Jagos, H., Oberzaucher, J.: Development of a Wearable Measurement System to Identify Characteristics in Human Gait – eSHOE. In: Miesenberger, K., Klaus, J., Zagler, W.L., Karshmer, A.I. (eds.) ICCHP 2008. LNCS, vol. 5105, pp. 1301–1304. Springer, Heidelberg (2008)

Author Index

- Abu Doush, Iyad I-108, I-376
Akiyama, Joji II-534
Al-Abdulkarim, Latifa II-1
Alagöz, Firat II-584
Albertin, Eugen II-568
Al-Hamdani, Sermed II-275
Al-Humaimedy, Abeer II-1
Al-Khalifa, Hend S. II-213
Al-Khonin, Manal II-33
Al-Khudair, Sara II-1
Allegra, Mario II-144
Älli, Sami I-544
Al Mahmud, Abdullah I-332, II-45
Alm, Norman II-576
Alonso, Fernando I-400, I-417
Al-Sarrani, Rawan II-33
Al-Shabanat, Hadeel II-33
Al-Sheaha, Maha II-25
Al-Wabil, Areej II-1, II-25, II-33
Al-Yahya, Maha II-1
Anaya, Pablo II-98
Andrade-Arechiga, Maria II-396
Aomatsu, Toshiaki I-116
Appell, Jens-E. II-568
Aranda, Joan I-312
Arató, Andrés II-247
Archambault, Dominique I-224
Arditi, Aries II-255, II-263
Atkinson, Matthew Tylee I-162
- Badr, Georges II-104
Baljko, Melanie I-605
Ballieu Christensen, Lars I-57
Ball, Simon I-74, I-433, I-665
Barberà, Ricard II-13
Barberis, Davide II-193
Barbieri, Thimoty I-324
Barrera, Carmen II-5
Bártek, Luděk I-358
Bates, Enda II-407
Batusic, Mario I-404, I-476
Becker, Josef I-273
Beckerle, Philipp II-128
Belda, Juan-Manuel II-13
- Bergmann, Stephan I-657
Bernabeu-Soler, Pablo II-136, II-140
Bernareggi, Cristian II-389
Bernier, Alex I-20
Bertoni, Giovanni II-423
Besio, Serenella I-235, I-511
Betke, Margrit I-139
Bianchi, Antonio I-324
Bigham, Jeffrey P. II-346
Böcker, Martin I-135
Bognár, Gábor II-247
Bohman, Ulla I-634
Bohner-Degrell, Claudia I-273
Borsci, Simone I-351
Bortolazzi, Enrico II-423
Bothe, Hans-Heinrich I-621, II-275
Boticario, Jesus G. I-484, II-5
Bou Issa, Youssef I-437
Boulay, Denis I-408
Brown, David J. I-227
Bühler, Christian I-273, I-550, I-569
Bulatov, Vladimir I-53
Bulgheroni, Maria I-235
Bumbalek, Zdenek II-160
Burger, Dominique I-20, I-408
Burzagli, Laura I-343
Butkute, Vilma I-227
Buzzi, Maria Claudia I-92
Buzzi, Marina I-92
- Cabrera, Rafael II-98
Calero Valdez, André II-584
Camarena, Javier I-320
Caprino, Francesca I-235, I-511
Carriço, Luís I-561
Casals, Alicia I-312
Casamitjana, Miquel I-312
Casas, Roberto II-5
Chandler, Andrew I-433, I-665
Chen, Ming-Chung I-193, II-37, II-120
Chiazzese, Giuseppe II-144
Chifari, Antonella II-144
Choi, Kup-Sze II-77
Chow, Chum-Ming II-77

- Chu, Chi Nung I-464, I-661
 Cockton, Gilbert I-216
 Como, Andrea I-343
 Constantinescu, Angela I-89, II-439
 Corrêa, Fernanda Cristina II-61
 Coughlan, James M. II-229, II-290
 Crombie, David I-1
 Cyrus, Michael II-304
 Czyżewski, Andrzej II-373
- Dale, Øystein I-300
 Darvishy, Alireza I-47
 de Beer, Roel I-332
 Debevc, Matjaž II-221
 De Boeck, Joan II-21
 Dedini, Franco Giuseppe II-61
 Dekelver, Jan II-21
 Delgado Neto, Geraldo Gonçalves II-61
 Delgado, Vivianne V. II-61
 de Witte, Luc I-537
 Dorigo, Martin I-47
 Duchateau, Sylvie I-408
 Dürnegger, Bernhard I-27
- Edler, Cordula I-630, I-649
 Edlinger, Günter II-85
 Eibegger, Mandana I-468
 El Ghouli, Oussama II-169
 Engelen, Jan I-5, I-69, I-81, I-200,
 I-550, I-553, I-569
 Erle, Markus I-657
 Ertl, Thomas I-100, II-506
 Evans, Shirley I-74
 Evreinova, Tatiana G. II-282
 Evreinov, Grigori II-282
- Fajardo-Flores, Silvia II-396
 Federici, Stefano I-351
 Feilmayr, Christina I-27
 Fels, Deborah I. I-247, II-576
 Felzer, Torsten II-128
 Ferreras, Alberto II-13
 Finat, Cecile II-5
 Fitzpatrick, Dónal I-12, II-407
 Flores-Barriga, Alfonso II-396
 Flores, Héctor II-296
 Formenti, Luciana II-423
 Fraternali, Piero I-324
 Frees, Bert I-81
 Fuertes, José L. I-400, I-417, I-503
- Fugger, Erwin I-288
 Fujiyoshi, Akio I-116, II-541
 Fujiyoshi, Mamoru I-116, II-515, II-523,
 II-541, II-548
 Fuwa, Teruhiko II-338
- Gaal, Andrea I-580
 Gabbanini, Francesco I-343
 Galinski, Christian I-573
 Garazzino, Nicola II-193
 García, Nuria II-13
 Gardner, Carolyn K. I-53
 Gardner, John A. I-53
 Gebert, Hermann I-621
 Gelderblom, Gert Jan I-537
 George, Jennifer I-216, II-69
 Gnanayutham, Paul II-69
 Goetze, Stefan II-568
 Gómez, Isabel I-147, II-98
 González, Ángel L. I-400, I-417, I-503
 Gowans, Gary II-576
 Gravenhorst, Claus I-35
 Greene, Cara Nicole I-495
 Greenough, Thomas I-216
 Grünbacher, Engelbert II-85
 Guerrero, Luis II-296
 Guger, Christoph II-85
 Gutiérrez y Restrepo, Emmanuelle II-5
- Hamidi, Foad I-605
 Hanafusa, Akihiko II-338
 Harakawa, Tetsumi II-443
 Hartjes, Rudolf I-449
 Hasegawa, Sadao II-443
 Hashimoto, Masami I-255, II-377
 Hashino, Satoshi II-235
 Heck, Helmut I-273
 Heimgärtner, Rüdiger I-613
 Hengstberger, Barbara I-404
 Hernáez, Inmaculada I-320
 Herstad, Jo I-208
 Hilzensauer, Marlene II-185
 Hlauschek, Walter II-600
 Hofmann, Philipp II-57, II-91
 Holone, Harald I-208
 Holzinger, Andreas II-221, II-556,
 II-584
 Horvath, Alexander I-47
 Hosono, Naotsune I-308, II-200
 Hossain, Mahjabeen Khaled I-527

- Hüttenrauch, Helge I-135
Hutter, Hans-Peter I-47
- Ienaga, Takafumi II-239
Ifukube, Tohru II-367
Iglesias, Ana I-591
Igual, Jorge II-136, II-140
Ikeda, Tomozumi II-338
Ikegami, Yuji II-515, II-523,
II-541, II-548
Imai, Atsushi I-62, II-367
Inoue, Hiromitsu II-200
Ito, Kiyohide II-447
Itoh, Kazunori I-255, II-377
Ivanchenko, Volodymyr II-229
Iwahana, Yukio I-62
- Jaballah, Kabil II-205
Jagos, Harald II-600
Javanovic, Rados II-49
Jaworek, Gerhard I-580
Jemni, Mohamed II-169, II-205
Jeschke, Sabina I-336
Jiménez, Javier I-591
Jones, Blake I-53
Jones, Elizabeth I-53
Juhász, Zoltán II-247
Jürgensen, Helmut II-498
- Kacorri, Hernisa I-178
Kagiyama, Yasuhiro II-447
Kakuta, Mari I-308
Kanahori, Toshihiro I-366
Kaneda, Naoki I-255
Kaneke, Takeshi II-515, II-523,
II-541, II-548
Karshmer, Arthur I. II-459
Katoh, Hiroshi II-271
Kawahara, Masaji II-308
Kawakami, Hiroshi II-447
Kayama, Mizue I-255
Keijers, Roy I-332
Keith, Suzette I-583
Kelly, Robert A. I-53
Kencl, Lukas II-160
Késmárki, László II-247
Kimuro, Yoshihiko II-239
Kipke, Siegfried I-131
Klaus, Joachim I-580
Ko, Chien-Chuan I-193
- Kobayashi, Makoto II-271, II-332
Kochanek, Dirk II-385
Kodama, Hiroshi II-377
Köhlmann, Wiebke II-498
Komori, Tomoyasu II-560
Kopeček, Ivan I-124
Koppenol, Tom II-45
Korn, Peter I-186
Kosec, Primož II-221
Kosikowski, Lukasz II-373
Kouroupetroglou, Georgios I-178
Kraus, Michael II-474
Krishna, Sreekar II-316
Krupiński, Robert I-154
Kummer, Franz I-35
Kurozumi, Koichi II-560
- Laudanna, Elena I-235, I-511
Lazaro-Flores, Juan II-396
Leitner, Marie-Luise I-392
Lepicard, Guillaume II-592
Leporini, Barbara I-92
Lewis, Jacqueline I-227
Li, Yunqiu I-162
Lin, Yun-Lung I-193, II-120
Liotta, Giuseppe I-351
Livingston, Nigel I-605
Linares, Raul II-136, II-140
Lo, King-Hung II-77
Lopes, Rui I-561
Lozano, Héctor I-320
Lunde, Frank II-304
- Machin, Colin H.C. I-162
MacKenzie, Ian Scott II-49, II-128
Mader, Christoph I-280
Magee, John I-139
Mariscal, Gonzalo I-503
Márkus, Norbert II-247
Marmol, Bruno I-20
Martens, Jean-Bernard II-45
Martín, Enric X. I-312
Martínez, Loïc I-400, I-417
Masuda, Fumihiko I-189
Matausch, Kerstin I-263, I-266, I-641
Matsuoka, Atsushi II-523
Mazurek, Przemysław I-154
Mazzone, Emanuela II-5
McMullen, Declan I-12
McNaught, Alistair I-74, I-433

- Mehigan, Tracey J. I-170
 Mele, Maria Laura I-351
 Merino, Manuel I-147, II-98
 Merlin, Bruno II-112
 Merlo, Gianluca II-144
 Miesenberger, Klaus I-35, I-43, I-243,
 I-266, I-404, I-531
 Miki, Yuuri II-338
 Millard, Emma I-665
 Minagawa, Hiroki II-152
 Minatani, Kazunori II-534
 Miró-Borrás, Julio II-136, II-140
 Miyagi, Manabi II-534
 Miyakawa, Masahiro II-308
 Miyamoto, Osamu I-384
 Miyaoka, Seiji II-377
 Mojahid, Mustapha I-437
 Molina, Alberto J. I-147, II-98
 Moranchel, Javier II-5
 Morandell, Martin M. I-288
 Moreno, Lourdes I-591
 Mori, Giulio I-92
 Mosies, Charlotte I-569
 Mubin, Omar I-332
 Müller, Karin I-89, II-439
 Murai, Yasuyuki II-308
 Murakawa, Kazuhiro II-560
 Murase, Haruka II-451
 Murray, Kyle II-346
- Nadig, Oliver II-482
 Nagar, Atulya K. II-403
 Nagasaka, Daisuke I-255, II-377
 Nagashima, Yuji I-308, II-200
 Nakazono, Kaoru I-308
 Navas, Eva I-320
 Nedbal, Dietmar I-456
 Neff, Flaithri I-170
 Nes Begnum, Miriam Eileen I-292
 Neubarth, Friedrich II-361
 Nicotra, Giuseppe II-423
 Nietzio, Annika I-425, I-468
 Nishimoto, Takuya I-368
 Nishizawa, Tatuso II-377
- Oberzaucher, Johannes II-600
 O Connor, Joshue I-413, I-444
 Ogawa, Tohru II-515
 Ohnishi, Noboru II-152
 Ohta, Hajime II-152
- Ohtsuka, Satoshi II-443
 Oikonomou, Theofanis I-186
 Olsen, Morten Goodwin I-425, I-468
 Oouchi, Susumu II-515, II-523, II-527,
 II-534, II-541, II-548
 Oriola, Bernard I-437
 Ortner, Rupert II-85
 Ossmann, Roland I-243
 Otani, Makoto I-255
 Othman, Achraf II-169
 Ottaviano, Simona II-144
 Ošlejšek, Radek I-124
 Övermark, Timo I-544
- Paap, Ken II-459
 Panchanathan, Sethuraman II-316
 Peböck, Birgit I-630, I-641
 Penichet, Victor M.R. I-92
 Petrie, Helen I-484, I-486, I-519
 Petz, Andrea I-531
 Petz, Gerald I-456
 Piccolo, Elio II-193
 Picón, Artzai I-320
 Pino, Alexandros I-178
 Pitt, Ian I-170
 Plhák, Jaromír II-354
 Pluke, Michael I-135
 Pontelli, Enrico I-108, I-376
 Pösl, Josef II-57, II-91
 Poveda, Rakel II-13
 Power, Christopher I-484, I-486, I-519
 Prescher, Denise II-482, II-506
 Pressl, Bettina I-280
 Prinetto, Paolo II-193
 Prüeckl, Robert II-85
 Pucher, Michael II-361
 Pühretmair, Franz I-263
- Radek, Christian I-273
 Raisamo, Roope II-282
 Raschke, Michael II-506
 Raynal, Mathieu II-104, II-112
 Rennies, Jan II-568
 Revuelta, Pablo I-591
 Rinderknecht, Stephan II-128
 Rivera, Octavio I-147, II-98
 Robinson, Thamburaj II-403
 Röcker, Carsten II-556
 Rodriguez-Ascaso, Alejandro I-135
 Roldán, Jose Ramón II-5

- Ruemer, Reinhard I-35, I-43
 Ruiz, Carlos I-503

 Sáenz, Mauricio II-296
 Sakharov, Vasily I-519
 Samperi, Valentina II-144
 Sánchez, Jaime II-296
 Santos, Olga C. II-5
 Sarantidou, Anna I-178
 Sasagawa, Chiharu II-338
 Sasaki, Nobuyuki II-443
 Schabus, Dietmar II-361
 Schiewe, Maria II-385, II-498
 Schilberg, Daniel I-336
 Schlegel, Thomas II-506
 Schmidt-Belz, Barbara I-583
 Schmidt, Michael II-490
 Schmitz, Bernhard I-100
 Schneider, Matthias I-135
 Schultz, Tanja I-595
 Secchi, Lorreta II-527
 Seiyama, Akitoshi II-447
 Seiyama, Nobumasa I-62, II-367
 Sekita, Iwao II-308
 Seta, Luciano II-144
 Sewell, John I-665
 Shahid, Suleman I-332
 Shen, Huiying II-229
 Shimada, Shigenobu II-451
 Shimizu, Michio I-255
 Shimizu, Yutaka II-451
 Shimojo, Makoto II-451
 Shiose, Takayuki II-447
 Shoda, Kiyotake II-560
 Sik Lanyi, Cecilia I-227
 Sillanpää, Niina I-544
 Silva, Ludmila Correa de Alkmin II-61
 Sloan, David I-162
 Snaprud, Mikael I-425, I-468
 Söllner, Matthias II-57, II-91
 Spalteholz, Leo I-605
 Spiewak, Vincent I-5
 Spiliotopoulos, Dimitris I-178
 Spindler, Martin II-474
 Standen, Penny I-227
 Steel, Emily I-537
 Steiner, Daniela I-476
 Strauss, Christine I-392, I-449
 Strobbe, Christophe I-5, I-81, I-569
 Sugimoto, Masahiko I-255

 Sugimura, Yukinobu II-239
 Suzuki, Masakazu I-53, II-415
 Swallow, David I-486, I-519

 Tacchella, Clarissa I-324
 Takagi, Tohru I-62, II-367, II-560
 Takatsuki, Akira II-447
 Takeuchi, Yoshinori II-152
 Taras, Christiane II-506
 Tatsumi, Hisayuki II-308
 Tazawa, Naoyuki I-62
 Tekin, Ender II-290
 Terada, Shuri II-338
 Teshima, Yoshinori II-513, II-515,
 II-523, II-541, II-548
 Tian, Yingli II-255, II-263, II-324
 Tiotto, Gabriele II-193
 Tito, Miguel II-13
 Toda, Kentaro II-447
 Totihara, Shinichi I-62
 Tzovaras, Dimitrios I-186

 Uchida, Yusuke II-451
 Unterfrauner, Elisabeth I-492
 Urra, Miguel II-13

 Valero, Marta II-13
 Van Hees, Kris I-200
 Van Isacker, Karel I-561, I-573
 Vannuffelen, Tim II-21
 Velasco, Carlos A. I-484
 Vieritz, Helmut I-336
 Vigouroux, Nadine I-437, II-592
 Vinagre, Manuel I-312
 Völkel, Thorsten II-385
 Votis, Konstantinos I-186
 Vy, Quoc V. I-247

 Wada, Chikamune I-189, II-239
 Wagner, Armin B. I-78
 Wakatsuki, Daisuke II-152
 Wald, Mike II-177
 Wallbruch, Rainer I-273
 Watanabe, Ken II-534
 Watanabe, Takayuki I-368
 Watanabe, Tetsuya II-534
 Watanabe, Yasunari II-515, II-523,
 II-541, II-548
 Watson, Sue I-433

- Weber, Gerhard II-431, II-466, II-474,
II-482, II-490, II-506
Weber, Harald I-649
Webster, Gemma II-576
Weiermair-Märki, Cäcilia I-492
Whitney, Gill I-583
Wieser, Manfred I-280
Wöß, Wolfram I-27
Wu, Chi-Fen II-37
Wu, Ting-Fang II-37
- Yamada, Sho II-235
Yamaguchi, Katsuhito I-53, II-415
Yamaguchi, Toshimitsu II-534
Yamamoto, Suguru II-451
- Yamazawa, Kenji II-515, II-523, II-527,
II-541, II-548
Yang, Xiaodong II-263
Yeh, Chih-Ching II-120
Yeh, Yao-Ming I-661, II-120
Yi, Chucai II-255
Yuan, Shuai II-324
- Zagler, Wolfgang II-600
Zelenka, Jan II-160
Zeng, Limin II-466
Zetterström, Erik I-135
Zieffle, Martina II-556, II-584
Zinke, Francis II-498
Zödl, Christian II-600