Springer Tracts in Advanced Robotics 111

Jean-Paul Laumond
Naoko Abe *Editors*

Dance Notations and Robot Motion

star

Springer

Springer Tracts in Advanced Robotics 111

Editors

Prof. Bruno Siciliano
Dipartimento di Ingegneria Elettrica
e Tecnologie dell'Informazione
Università degli Studi di Napoli
Federico II
Via Claudio 21, 80125 Napoli
Italy
E-mail: siciliano@unina.it

Prof. Oussama Khatib
Artificial Intelligence Laboratory
Department of Computer Science
Stanford University
Stanford, CA 94305-9010
USA
E-mail: khatib@cs.stanford.edu

More information about this series at http://www.springer.com/series/5208

STAR (Springer Tracts in Advanced Robotics) has been promoted under the auspices of EURON (European Robotics Research Network)

Jean-Paul Laumond · Naoko Abe
Editors

Dance Notations and Robot Motion

Springer

Editors
Jean-Paul Laumond
LAAS-CNRS
Toulouse Cedex 4
France

Naoko Abe
LAAS-CNRS
Toulouse Cedex 4
France

ISSN 1610-7438 ISSN 1610-742X (electronic)
Springer Tracts in Advanced Robotics
ISBN 978-3-319-25737-2 ISBN 978-3-319-25739-6 (eBook)
DOI 10.1007/978-3-319-25739-6

Library of Congress Control Number: 2015952769

Springer Cham Heidelberg New York Dordrecht London

Printed on acid-free paper

Springer International Publishing AG Switzerland is part of Springer Science+Business Media (www.springer.com)

Foreword

Robotics is undergoing a major transformation in scope and dimension. From a largely dominant industrial focus, robotics is rapidly expanding into human environments and vigorously engaged in its new challenges. Interacting, assisting, serving, and exploring with humans, the emerging robots will increasingly touch people and their lives.

Beyond its impact on physical robots, the body of knowledge robotics has produced is revealing a much wider range of applications reaching across diverse research areas and scientific disciplines, such as biomechanics, haptics, neurosciences, virtual simulation, animation, surgery, and sensor networks among others. In return, the challenges of the new emerging areas are proving an abundant source of stimulation and insights for the field of robotics. It is indeed at the intersection of disciplines that the most striking advances happen.

The *Springer Tracts in Advanced Robotics* (*STAR*) is devoted to bringing to the research community the latest advances in the robotics field on the basis of their significance and quality. Through a wide and timely dissemination of critical research developments in robotics, our objective with this series is to promote more exchanges and collaborations among the researchers in the community and contribute to further advancements in this rapidly growing field.

The edited volume by Jean-Paul Laumond and Naoko Abe is focused on human motion analysis within a unique artistic–scientific multidisciplinary context embracing robotics, computer science, neuroscience, dance notation systems, and choreography. The twenty chapter collection originates from the workshop “Dance Notations and Robot Motion” held at LAAS–CNRS in Toulouse in November 2014, ranging from motion computation to segmentation, from robot programming to gestural languages, from behavioral objects to robot interaction, and from abstract symbol manipulation to physical signal processing. The results described in the volume provide a number of useful tools toward better understanding the computational foundations of anthropomorphic action.

Rich of heterogeneous contributions by robotics researchers and dance experts in the field, this volume constitutes a very fine addition to STAR!

Naples, Italy
August 2015

Bruno Siciliano
STAR Editor

Preface

A motion is perceived by others from its completion into the physical space. At the same time, the same motion originates within the personal body space. Identifying what is happening in the body when a person is moving is not an easy task. Human motion analysis focuses on the relationship between physical space and body space. How to bridge both physical and body spaces? As biomechanists or neuroscientists, roboticists face the same but differently worded question: For roboticists, the physical space is the task space or the operational space in which robot actions are expressed; the body space is the control space or the configuration space according to the considered robotic system. The question is related to robot programming, i.e., to the competition between abstract symbol manipulation and physical signal processing.

In the dance field, tools for analyzing and transcribing human movements, i.e., the so-called dance notation systems, have been developed to enhance dancer performance. The main purpose of dance notations is to store choreographic works and knowledge of dance techniques by translating movements into specific ways as abstract symbols, letters, abbreviations, musical notations, stick figures, etc. In Western culture, there are almost 90 dance notation systems, from the first appearance in fifteeenth century to the present. Among the currently used ones, we find the Kinetography Laban, the Benesh Movement Notation, and the Eshkol-Wachman Movement Notation.

In this context, it is natural to gather roboticists, computer scientists, neuroscientists, dance notation system researchers, and choreographers, in order to promote a multidisciplinary research on human motion analysis. This was the objective of the workshop "Dance Notations and Robot Motion" held at LAAS–CNRS in Toulouse in November 2014.[1] How an anthropomorphic action is decomposed into a sequence of motions? How an emotional state is reflected in a motion? How to

[1]The workshop took place in the framework of the Anthropomorphic Motion Factory launched by the European Project ERC-ADG 340050 Actanthrope devoted to exploring the computational foundations of anthropomorphic action.

describe a dance in terms of a sequence of elementary motions? Such questions and many others were the ingredients for stimulating discussions. The first challenge of the meeting was to reach a mutual understanding allowing a choreographer to access robotics concepts, or a computer scientist to understand the subtleties of dance notation.

This book intends to keep traces of this unique meeting. It results from the willingness of authors to share their own experiences with others.

The reader is then introduced to the basics of dance notations. At the same time, it can delve into chapters reporting dancing robot performances. Computational issues in motion notation are introduced via the dual perspective of both automated motion scoring and dance scoring-based motion generation. Motion segmentation appears as a way to better program robots. Cognitive dimensions are reflected in chapters dealing with gestural languages, behavioral objects, or robot interaction. The reader will recognize the multidisciplinary character of the research through the heterogeneous style of the textbook.

We thank all the authors for their effort in making their own research field accessible to others. We thank also the reviewers and the editorial committee (Marion Bastien, Antonio Camurri, Henner Drewes, Kozaburo Hachimura, Dana Kulic, Shelly Saint-Smith, and Gentiane Venture) for their help in editing the book.

Toulouse
July 2015

Jean-Paul Laumond
Naoko Abe

Contents

Towards Behavioral Objects: A Twofold Approach for a System of Notation to Design and Implement Behaviors in Non-anthropomorphic Robotic Artifacts

Samuel Bianchini, Florent Levillain, Armando Menicacci, Emanuele Quinz and Elisabetta Zibetti

Abstract Among robots, non-anthropomorphic robotic artifacts are in an interesting position: the fact that they do not resemble living beings, yet impart a sense of agency through the way they move, motivates to consider motion as a source of expressivity in itself, independently of any morphological cues. This problematic is considered in parallel to the question of movement notation and the different levels

---

This publication is part of a research program ("The Behavior of Things") that is supported, through the Labex Arts-H2H, by Investissements d'Avenir (ANR-10-LABX-80-01) of The French National Research Agency (ANR).

S. Bianchini (✉)
The École nationale supérieure des Arts Décoratifs, Ensadlab/Reflective Interaction,
31 rue d'Ulm, 75240 Paris Cedex 05, France
e-mail: samuel.bianchini@ensad.fr

F. Levillain · E. Quinz · E. Zibetti
University of Paris 8 Vincennes-Saint-Denis, 2 rue de La Liberté,
93526 Saint-Denis Cedex, France
e-mail: florent.levillain@userlab.fr

E. Quinz
e-mail: equinz@univ-paris8.fr

E. Zibetti
e-mail: ezibetti@univ-paris8.fr

F. Levillain · E. Zibetti
The École pratique des hautes études (EPHE), 4-14 rue Ferrus, 75014 Paris, France

F. Levillain · E. Zibetti
The Cité des sciences et de l'industrie, 30 Avenue Corentin Cariou, 75019 Paris, France

A. Menicacci
Dance Department, The Université du Québec à Montréal (UQAM), Case Postale 8888,
succursale Centre-ville, Montréal (Québec) H3C 3P8, Canada
e-mail: menicacci.armando@uqam.ca

E. Quinz
The École nationale supérieure des Arts Décoratifs, Ensadlab-Reflective Interaction,
31 rue D'Ulm, 75240 Paris Cedex 05, France

J.-P. Laumond and N. Abe (eds.), *Dance Notations and Robot Motion*,
Springer Tracts in Advanced Robotics 111,
DOI 10.1007/978-3-319-25739-6_1

of abstraction that one may consider when reflecting on movement and its relation to a spatial, temporal and social context. This is through a twofold perspective, drawing on both dance notation and cognitive psychology, that we consider the question of movement notation, and its relation to expressive gestures and psychological attributes. To progress in the direction of a system of notation that could integrate the qualitative, relational, and behavioral aspects of movement, we propose different typologies and a model of constraints to analyze, conceive and implement behaviors in robotic artifacts.

1 Introduction: Analyzing, Designing, and Implementing Behaviors for Non-anthropomorphic Objects?

In light of new forms of domestic and social robotics, many objects in our daily lives could be endowed with the capacity to move autonomously, to act and react, even to adapt flexibly to their environment. These robotic artifacts would be able to carry out actions that, though partially scripted, could give the impression of being motivated by and executed in pursuit of goals, and might even be considered intelligent [5, 6]. Because these objects would not resemble humans or other animal forms, their expressive potential would emerge mostly from dynamic and contextual cues: the way they move, the way they engage with the environment, the way they react to social agents.[1] From this behavioral specificity arise two interesting questions. The first is related to the "disembodied" aspect of their movement: if we consider motion to be a source of expressivity in itself, we need to find a way to extract expressive movements that apply as much as possible independently of the physical structure of the object. The second question has to do with the way human observers spontaneously interpret motion: to convey emotions and other psychological attributes, we ought to determine which motion cues need to be considered, and how those cues interact with the context of a particular behavior.

How can we annotate the movement of robotic artifacts, and how can we implement expressive behaviors in these non-anthropomorphic (and non-zoomorphic) structures? How can we dispense with the human body as a system of reference, while still maintaining a movement quality, suggestive enough to recall a living being, even to evoke a personality? These questions are approached within a project both practical and theoretical, bringing together several disciplines. It is through a twofold perspective, drawing on both dance notation and cognitive psychology, that we consider possible responses to the question of movement notation at different levels of abstraction and related to psychological properties. Our project has only just gotten

[1]We call "behavioral objects" the moving artifacts that a human observer perceives as acting in a meaningful way. Those artifacts need not necessarily be robotic, although their "behavioral" potential is of course reinforced by their ability to react to certain events. Cf. [4].

off the ground, but we have already conducted studies investigating autonomous robotic artworks [31] and organized workshops[2] based on the development of a modular robotic toolkit to quickly prototype robotic artifacts.[3]

Analyzing movements that are not located within a familiar system of reference, be it a human body or other corporeal property, is not without problems when considering existing dance notation systems. Furthermore, movement notation is not formalized enough to be integrated into a software environment, even though some works are pursuing that goal.[4] Yet offering the possibility of easily configuring the behavior of robotic objects is a stimulating perspective, both for artistic creation, design, design of interactions, and to introduce new practices aimed at informed users (see [25], for a similar perspective).

Working on the expressive quality of movements and their interpretation in terms of meaningful behavior requires the elaboration of original tools for evaluation and annotation. In pursuit of that goal, we propose a work comprised of three main sections:

1. from the viewpoint of the understanding of action in cognitive psychology, we delineate an articulation between movement and behavior through crucial psychological attributions;
2. from the viewpoint of dance and musical notation, we elaborate on the notion of gesture and interpretation to advance the description and formalization of basic movements, independent of the human body;
3. a comparison of these two approaches is attempted in order to build the first blocks of a system of notation devoted to the description, conception, and implementation of behavioral patterns in robotic artifacts.

[2]For example: "The Misbehavior of Animated Objects" Workshop, 8th International Conference on Tangible, Embedded and Embodied Interaction (TEI 2014). February 2014, Munich. URL, July 9, 2014: http://www.tei-conf.org/14/studios.php#s9.

[3]The *MisB* Toolkit developed by the EnsadLab/Reflective Interaction team, under the direction of Samuel Bianchini, by Didier Bouchon, Colin Bouvry, Cécile Bucher, Martin Gautron, Benoît Verjat, and Alexandre Saunier, in the context of the project The Behavior of Things, with the support of Labex Arts-H2H and the Bettencourt Schueller Foundation. For more information, URL, July 9, 2014: http://diip.ensadlab.fr/fr/projets/article/the-misb-kit.

[4]See for instance project LOL (Laban On Lisp) by Fred Voisin, from Myriam Gourfink's research in choreography. Cf.: http://www.fredvoisin.com/spip.php?article164 et Frédéric Voisin, *LOL: Un environnement expérimental de composition chorégraphique*, in *Ec/cart*, vol. 2, Eric Sadin ed., Dif'Pop, Paris, 2000.

1.1 The Understanding of Action: A Psychological Framework to Understand Relationships to a Non-anthropomorphic/Non-zoomorphic Robotic Artifact

Naive psychology is an ensemble of psychological faculties that collaborate to produce a rich and coherent picture of the social world around us. As human beings, we are endowed with refined capacities to perceive a vast range of attitudes in others creatures—humans mostly, but not only; to impart goals, desires and beliefs; to evaluate actions performed by other agents; to assess their motivation to cooperate; to determine their emotional state; or to engage in relevant interactions and communication behaviors. The core properties of naive psychology are the ability to extract relevant information from a flow of movement, which implies the possibility of segmenting this flow into meaningful units [36], and the ability to categorize actions in terms of whether the action is fortuitous or not, meaning whether it reflects an intention from an agent or not [1, 35, 44].

When observing a robotic artifact, we usually exploit these psychological resources to interpret the sequences of actions produced by that artifact, much as we would interpret the behavior of an animal or a human being. Anthropomorphism is a reflection of the "overuse" of our interpretative skills when confronted with an entity that does not necessarily possess the psychological attributes we impart to it [13]. We cannot help but attribute intentionality to robots' actions, and endow them with mental states [23, 50], and this process of anthropomorphization determines a type of interaction that is social in nature.

In the process of designing effective motions to express affects and aesthetic feelings, taking some basic principles of naive psychology into account may be of considerable help in informing the relationship between humans and a robotic artifact. Since the robot does not need to possess the psychological properties we want it to express, all that is needed are certain convincing cues—in the way the object moves and interacts with its surroundings—to trigger specific psychological attributions.

1.2 The Process of Psychological Attribution

In social psychology, the term *attribution* refers to the process of explaining behavior, as well as inferring traits from behavior [32]. The two meanings are related in the sense that ascribing a psychological trait is often a way to explain the reasons a person is doing something (e.g. *he goes to the door because he wants to go out; he wants to go out because he is bored*). The process of attribution has been seen as an attempt by the cognitive system to recover the causal and social structure of the world [44], or to extract invariants from an agent's stream of ongoing behavior [22]. Psychological attributes may be considered the output of a process that seeks to determine stable properties in the social world: intentions, motives, traits, and sentiments participate in the explanation of a current perceived action, while also forming a basis upon which to predict future behavior. Recent development in HRI have included this process of psychological attribution to elaborate credible and meaningful interactions with robots (e.g. [3, 15, 25]).

In the model we propose to qualify the psychological attributions given to a robotic artifact, the process of attribution evolves with the complexity of the robot's observable behavior: a human observer collects cues from the robot's movements and various transformations, and infers psychological properties about its perceptive skills (how much of the environment the robot is aware of), its ability to plan an action, to reason, to make decisions adapted to specific circumstances, etc. As the possibilities of movement and flexible behavior increase, so do the inferences and the resulting attributions.

What is interesting to consider in the process of conceiving such an artifact is how we want to limit the psychological attributions in a human observer. Do we want to give the impression of a conscious agent, willingly acting to accomplish specific goals? Or do we want to represent a more erratic, perhaps inscrutable behavior? These psychological traits and attitudes that a human observer spontaneously ascribes to a moving object depend on specific behavioral parameters. In the model we propose, these behavioral parameters are categorized according to three levels of interpretation: the animacy level, the agency level, and the social agency level (Fig. 1). In essence, the three levels correspond to the three following questions: Does the object look alive? Does the object appear to act intentionally? Does the object appear to interact socially with other agents?

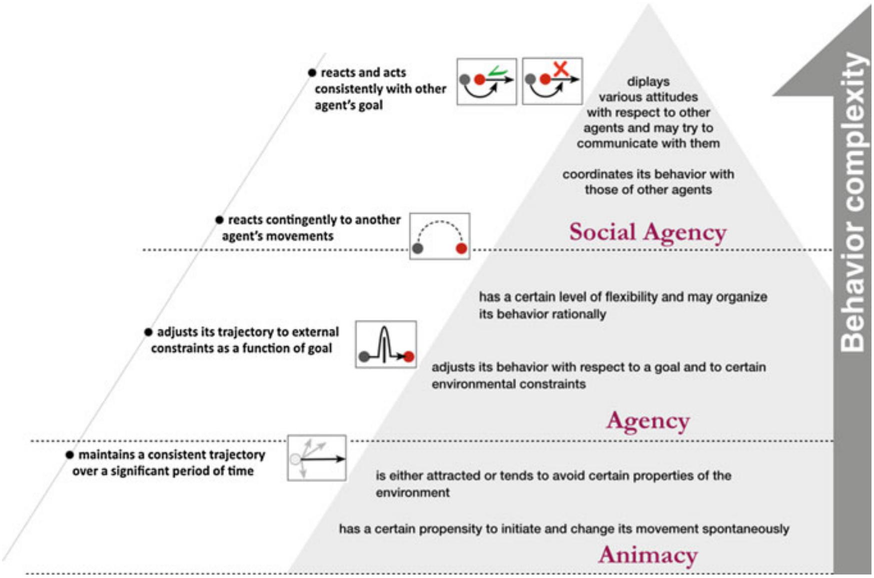


Fig. 1 Three levels of interpretation to describe behavioral patterns

At the **animacy level**, a moving object is perceived as basically autonomous: it seems to move on its own and may give the impression of reacting to its immediate environment. While the object may initiate an interaction with its surroundings, through movements of attraction or repulsion, its behavior is still too erratic to be qualified as fully intentional. At the **agency level**, the behavior of a moving object is construed as goal-oriented: the object seems to act for the realization of specific actions and to organize its behavior flexibly and rationally to reach those ends. At the **social agency level**, a moving object already identified as intentional is also granted some social skills: its behavior is not only related to the properties of the immediate environment, but also to the behavior of others agents with whom it may try to interact and communicate.

More specifically, objects situated at the animacy level have a propensity to initiate and change the direction and speed of their movements spontaneously. These simple motion cues have a strong impact on the naive perception of animacy [2, 41]. Seeing a simple dot on a screen changing direction suddenly is often enough to give the illusion of a living being, although the strength of the impression depends largely on the importance of the direction change and possibly of the speed change [47]. The transition from an autonomous yet erratic entity to a goal-oriented entity will depend on the consistency of the behavior over time: if the object maintains consistent trajectories rather than changing direction at every turn, we are likely to infer in it the ability to maintain consistent objectives [16]. This impression will be reinforced by the possibility of connecting the robot's movement to a specific location towards which it appears to be moving [17, 48]. Compared to the animacy level, entities at the agency level have a flexibility in their behavior; they act purposefully and may organize their actions in a rational manner so as to reach a goal in the most efficient way. Seeing a moving object adjusting its behavior with respect to external constraints is a potent indication that it does indeed possess the ability to select appropriate means to achieve a goal [12]. At the social agency level, agents possess the capacity to act not only relative to their own goals but also relative to other agents' goals. Based on the apparent coordination of two moving objects, an observer can determine whether their respective goals converge (and thus attribute a positive social attitude) or diverge (and attribute a negative social attitude) [21].

Depending on the cues offered by the object's behavior, and the level at which it is construed, certain personality traits may be spontaneously attributed to the object. A robotic artifact may look curious, mischievous, indifferent… based on the way it organizes its movements, reacts to external events, or interacts with other agents. An object seen merely as autonomous (animacy level) may not be granted any proper personality traits, its behavior simply not being organized enough to be qualified with a psychological attribute. However, as soon as the object's behavior can be related to intentions (agency level), a vast range of psychological dimensions are available to qualify it, for instance how effective the agent is in the accomplishment of an action, how rational, how persistent, how thorough it is in the exploration of its environment, etc. At the social agency level, even more psychological components can be used to qualify the agent's behavior, based for instance on the proficiency with which it interacts socially, its propensity to engage in communication behaviors, or its tendency to pursue positively or negatively valued behaviors.

1.3 Emotion and Empathy

The previous model is essentially concerned with the attribution of perceptive skills and abilities to adjust to environmental and social constraints. It does not tackle the affective component of an organism's behavior. However, the affective dimension is very important in human–robot interaction (e.g., [40]), because among many other functions it is strongly related to empathy. Research on empathic agents are divided in two main branches: agents that simulate empathic behavior towards the users and agents that foster empathic feelings from the users [38, 39]. While anthropomorphic robots are capable of facilitating social interactions with humans since they are utilising social interaction dynamics analogous to those observed in human-to-human interactions (e.g., [37]), some recent studies show that it is also possible to achieve these effects by using minimalistic behaviors to convey a sense of sociability from a non-anthropomorphic robot (e.g. interest, rudeness, empathy, etc. [43]).

Hence, to complete the description of the types of attributions a person may entertain when observing a robotic artifact, we shall evoke the relationship of empathy, defined as the ability to understand and respond to others' affective states [24, 42], that may take place during the interaction between a human and a robotic artifact (e.g., [14, 23]). When observing an animal or a human, we are sensitive to its affective state: we sympathize with the actions it tries to accomplish, we feel for its frustrated efforts, we experience an echo of its pain or pleasure. This affective relationship is also dependent on cues that may be expressed through a robotic artifact's motion (see [26]) for an extensive survey on body movements for affective expressions).

Very schematically, an observer may identify three different states an organism can find itself in: a state of comfort, a state of discomfort, or a state of distress. As depicted in Fig. 2, the **state of comfort** is the basic state. Despite a certain level of natural variations occurring in different physical variables (temperature, blood pressure, glucose level, oxygenation of blood, etc.), regulatory mechanisms conspire to maintain these variables at a safe level so that the organism is not lacking in any crucial life components (this is the principle of homeostasis, initially described by Claude Bernard and Walter Cannon). But, as the environmental conditions change constantly, the organism is under pressure from external events that may cause a homeostatic imbalance, corresponding to the **discomfort state**. Through negative feedback, corrective mechanisms launch and bring the system back toward normal variations (vasodilation, decreased metabolic rate, sweating, etc.). Negative feedback can be internal and unobservable, but it can also consist of an observable motor behavior, for instance an organism may look for shade in a heated environment.

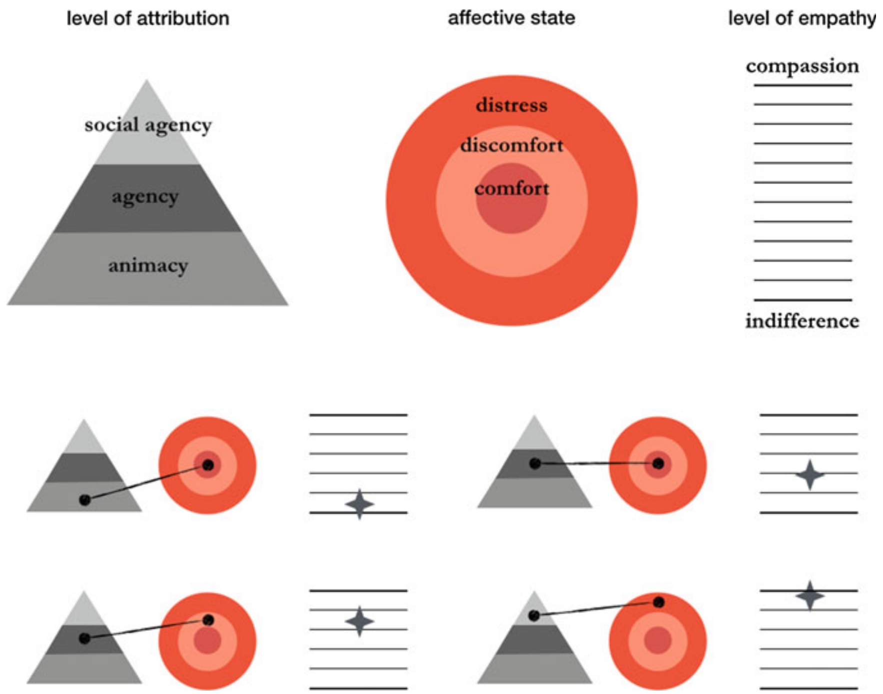


Fig. 2 Relationship between psychological attribution and affective states

Sometimes, external variations are so strong that the organism cannot adapt rapidly enough, causing temporary or permanent damage to some life functions and leading to a **state of distress**. If this does not lead to death, a repair process will take place, leading back to the discomfort zone where damage to the body is transformed into a mere uncomfortableness.

Behavior, emotion, and empathy are closely intertwined. For an organism, engaging in a behavior is one way to maintain stable life conditions and stay in a state of comfort, either reactively, by acting directly on the environment, or proactively, by foraging or removing potential threats. Emotions may accompany these activities when stable conditions are no longer guaranteed: stress may be the primary form of emotion, being a signal that indicates to the organism that it needs to react quickly to some external variations before any internal adaptation takes place. In certain species, emotions also constitute a social signal to congeners, telling them to prepare themselves for a threat. This is where empathy comes in. It can be understood as a response to the signals an organism displays when outside its comfort zone and seeking to return to normal conditions. Thus, the natural connection provided by empathy can be channeled when trying to modulate the expressivity of a robotic artifact, providing cues about the different states, from comfort to distress, in which it can find itself.

1.4 Relationship Between Attribution and Expressivity

From indifference to compassion or fear, the relationship to a robotic artifact may take many forms. Based on the models we described, we can imagine a rather systematic relationship between the psychological properties attributed to a robotic artifact and its expressivity, which we define as the ability to convey some affects and create a relationship of sympathy or empathy (Fig. 2). Two sets of behavioral cues determine the robot's expressivity (although those two sets can overlap): the ensemble of cues that provide some information regarding the perceptive and cognitive skills of the robots, and the ensemble of cues that indicate in which state of comfort/discomfort the robot finds itself. Manipulating these cues may make it possible to vary the relationship between a human observer and the robot, and to create different types of interactions, from simple coexistence to active communicative interaction.

For instance, a robot construed as merely autonomous, without any cues about intentional behavior, may evoke nothing other than relative indifference. But adding some evidence that it is in pain would conjure a degree of empathy. That empathy could be strengthened by cues indicating the robot's ability to pursue intentional actions, provoking sympathy (a tendency to put oneself in others' shoes) for the action in which the robot is engaged. Adding a social component to the robot's behavior allows for some additional values to be considered, especially by determining how positively or negatively it relates to one's own intentions and goals.

The model we propose to describe some general properties of psychological attribution and the motion cues on which it is based makes it possible to elaborate some general objectives regarding the expressivity of a robotic artifact. Before any concrete implementation, one could determine what kind of behavioral properties to impart to the robot in order to elicit a psychological attribution at a particular level of interpretation. This "psychological prototyping" is merely a starting point in the process of determining the movement qualities a robot should possess to attain an appropriate level of expressivity. It should therefore be supplemented with tools designed to precisely qualify the nature of the movement we want to implement in a robotic artifact, both in its intrinsic quality and its temporal execution.

2 Towards an Analysis of Qualia of Gesture: Gesture and Meaning

2.1 Take Note of Movement

If, to paraphrase Goodman [19], a score is the object that guarantees the identity of a work allowing for different incarnations, interpretations, and performances, we should consider the need to invent a system of notation in robotics. Should we speak rather of description or transcription or an analytical grid for analysis? To

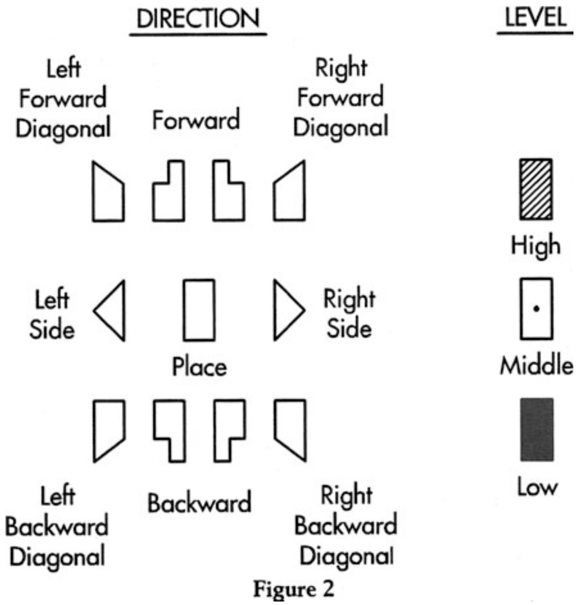


Fig. 3 Eight directions in each spatial plane and three levels of direction for body segments

record human movement, one of the most effective and widespread systems is Kinetography Laban or Labanotation (Fig. 3). In that system of notation, for example, there are only 8 directions in each spatial plane, and three levels of direction for body segments (high, middle, and low): a system of notation must necessarily reduce the contingency and complexity of a movement to a reproducible, representative system that is sufficiently precise to account for a phenomenon, but ambiguous enough to be able to describe a similar, but different, behavior.

To create a behavior notation or transcription system, it is essential to start from geometry. Movements and rotations on the three planes must be noted for all segments and in relationship to a "root" of all the segments. Since the human body is radially symmetric in kinetography, the basic orientation of the figure is the pelvis. But if we want to transcribe or note a behavior, it is necessary to do more than describe the relative motion of the parts of the body performing it, especially when discussing the behavior of non-anthropomorphic robotic figures. To understand why and how to note the other aspects of a behavior, we must start by differentiating movement from gesture. A movement is the shifting of an object or one of its segments in space and in time relative to a system of reference. An apple that falls from a tree, a train hurtling at top speed, or a pendulum perform movements, but not gestures.

2.2 From Movement to Gesture: Performative Agogics

The definition of the gesture develops around the question of meaning, the signifié attributed to a movement (by the author and/or the observer): a gesture is a movement that has a specific meaning within a specific culture. That meaning is not necessarily a matter of a coded semantics—modern and contemporary dance emerged, after all, in opposition to a codification of both bodies and movements. If movement is denoted by its form, gesture appears as an element of a discourse. But how is that discourse defined? And how can the expressiveness of a gesture be defined?

Taking the example of a simple gesture (extending a forearm previously folded back over the humerus held out forward), Hubert Godard explains that the gesture can of course be accomplished in a number of ways, but that it will almost always be understood through the application of two pure trends: on the one hand, it can be seen as an order or a command, on the other as a request or an appeal. The two movements, however, should, be noted exactly in the same way in Kinetography (extension of the forearm with the arm oriented on the sagittal plane facing forward, middle level). To summarize, for Hubert Godard, the reading of the meaning of a gesture is not limited to the analysis of the motions of the moving parts, but is more a relationship to postural tensions that act like a background for the frontal figure, which corresponds to what medicine defines as *Anticipatory Postural Adjustments* (APA).[5]

Hubert Godard proposes calling the ensemble of that activity "pre-movement" [18]. Pre-movement is based on a specific relationship to weight and gravity, and precedes and anticipates movement. The same gestural form—such as an arabesque —can be charged with different meanings depending on the quality of the pre-movement, which undergoes very large variations as the form endures. It is what determines the tautness of the body and defines the quality, the specific color of each gesture. Pre-movement acts upon gravitational organization, meaning the way the subject organizes its posture in order to remain erect and respond to the law of gravity in that position. A whole system of anti-gravity muscles, whose action largely escapes vigilant consciousness and will, is responsible for our posture; they are the ones that maintain our balance and allow us to remain standing without having to think about it. It so happens that these muscles are also the ones that register our changes of affective and emotional states. Thus, any change in our posture will have an effect on our emotional state, and conversely, any affective change will lead to a modification, even an imperceptible one, of our posture.

The analysis of "pre-movement" is that of the non-conscious language of posture. Erect posture, beyond the mechanical problem of locomotion, already contains

[5]The literature on APA is vast, but among the first to have had the intuition and then to have conducted experimental research were: N.A. Bernstein, *The coordination and regulation of movements*, Oxford, Pergamon, 1967; S.M. Bouisset, M. Zattara, "A sequence of postural movements precedes voluntary movement", *Neuroscience Letter*, 1981, n. 22 pp. 263–270.

psychological and expressive elements, even prior to any intentionality of movement or expression. It is pre-movement, invisible and imperceptible for the subject itself, that activates at the same time the mechanical and affective levels of its organization. Depending on our mood and the imagination of the moment, the tensing of the calf, which prepares without our knowledge the movement of the arm, will be stronger or weaker, and will therefore change the perceived significance. The culture and history of a person, and their way of experiencing and interpreting a situation, will induce a “postural musicality” that will accompany or catch out the intentional gestures executed [18].

So, to understand and read a gesture, alongside what happens in the geometry of the kinesphere (the sphere of motions in the directions of space), it is necessary to analyze the organization of the whole gravitational relationship and the tensions that anticipate and underlie the entire equilibrium. Those are what produce voluntary movement within the kinesphere. To summarize, a gesture is, according to Hubert Godard, a movement plus a pre-movement.[6]

United in gesture, movement and pre-movement produce a unique dynamic that Laban called the “dynamosphere” or the “connection between the outer movement path and the mover’s inner attitude” [27], meaning all of the APAs, plus the energy used (the measurable product of the work) and the play of co-contractions (or non co-contractions) of agonistic and antagonistic muscular groups. To note those dynamics, Laban invented a notational system less well-known than his kinetography, called the *Effort Shape* [28]. The system only integrates four parameters, and does not allow for the notation of the shape of a movement, but does make it possible to note all of the fundamental dynamics (Fig. 4).

In general, perception of dynamics is visual, but can also be auditory. Hubert Godard often presents this example: someone lives in a house they know well. The house has a wooden staircase that makes noise when someone goes up or down the stairs. If the person knows the other inhabitants of the house well, just by listening to the sounds made by the stairs when someone uses them, he or she can know who is going upstairs and that person’s tonic and psychological state. Godard’s example is interesting because it evokes a listening context, and therefore refers to the sound effect of body movements. The listening context described is that of acousmatics—which designates the fact of hearing sounds without seeing their source. Launched

---

[6]The experiment carried out in the framework of the TechLab workshop of the Dance Monaco Forum (by a team that included, among others, H. Godard, A. Menicacci, E. Quinz, the research team in IRCAM directed by F. Bevilacqua, and the research scientist in behavioral neuroscience I. Viaud-Delmon, CNRS) led to an interesting discovery. By using optical fiber flexion sensors to measure the lateralized movements of two dancers, in particular making repeated arm movements to the right and left, it was possible to determine that the movements involved a whole series of adjustments that were not bilaterally symmetrical as we might have imagined. That indicates that the geometric kinesphere is only a theoretical premise and that when it is experienced through the contingency of a gesture it is dented and asymmetrical. The perceived space is not homogeneous and it contains variable directions and intensities. Cf. A. Menicacci, E. Quinz, “Etendre la perception? Biofeedback et transfert intermodaux en danse”, in “Scientifiquement danse”, *Nouvelles de danse 53*, Brussels, 2005, pp. 76–96.

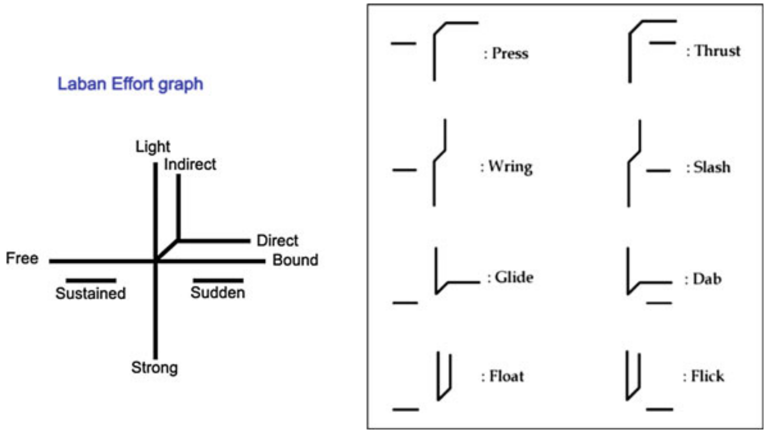


Fig. 4 The Laban Effort graph, and the eight fundamental dynamics

by Pythagoras to make his disciples concentrate on his voice, the technique was taken up by composers of musique concrète in the 1950s and 1960s. In that context, several attempts were made to analyze and note down the qualities of what Pierre Schaeffer called "acoustic objects" [11, 45, 46], objects that escaped—through their irregular morphology—traditional systems of notation. Sachaeffer, in the propositions in his *Solphège de l'objet sonore* (Music Theory of Acoustic Objects), explored several approaches to notation for the timbre parameter—offering notions such as the *harmonic timbre* (which defines the constellation of harmonics that surround a sound), the *grain* (texture of the sound matter), or the *allure* (oscillation of the ensemble of the sound's characteristics). The sound parameter that could prove most promising for our research is agogics.

In music, agogics designates the slight modifications of rhythm that affect the interpretation of a piece of music in a transitory way, as opposed to a strict and mechanical performance. If timbre concerns the instrument, agogics concerns the player. It manifests itself in the form of slowing, acceleration, caesurae, or pauses that "color" and personalize the performance of a movement, like identifying traits or hesitations. Agogics can be useful in our research because it makes it possible to re-incorporate qualia—which are characteristic of gesture—into movement.

There have been several attempts to notate agogics: episemes (from the Greek ἐπίσημον, "distinguishing feature"), which specify a nuance of rhythm—especially in Gregorian chant, up through diagrams supplied by the analysis of physical energy of the performer's gesture [29, 30]. Even if agogics only functions on one level, that of the distribution of accents (basically, energy in movement) and is reflected in the choice of emphasis in one plane of space, it can make it possible to work on the idea of an "individualizing" effect that manifests itself in the form of a script that adds bumps to the linear and fluid (therefore mechanical) performance of a movement that can give it "gestural" connotations: slowing down, accelerating, backing up, etc. Agogics is interesting because, like gesture, it lies at the crossroads between the physical (anatomy) and the psychological.[7] Similarly, Hubert Godard's analysis of the kinesphere is interesting in that it connects the physical to the psychological, considering that a subject's history can also be read through physical elements.

2.3 From Gesture to Behavior

Once the dynamic aspect has been integrated into the geometry and temporality of movements, only one parameter remains to read movement in humans: the relationship to the context. A gesture is always performed in a historically and geographically determined context, one that is also read, interpreted, and categorized by the moving subject. It is the analysis of what happens in the kinesphere, plus what happens in the dynamosphere, plus elements contextualizing the gesture that make possible the overall reading of the meaning of the gesture, or what Hubert Godard calls the "gestosphere",[8] or all of the emotional and symbolic data that connect the person making the gesture to their environment.

Behavior can be a series of predetermined (such as a fixed choreography) or generative actions, a set of rules that determine a sequence of decisions and actions in time which can be not completely frozen in their structure (such as an improvisational system in dance or music). Thus defined, even the motions of a sensorless robot can be described as a behavior, because even the most rigid and purely functional movements have their own precise spatial qualities established by a choreographed program and by a relationship to gravity. The simplest robot therefore has a certain type of behavioral relationship to its environment that is its own. That behavior, however, is of an autistic type: the robot does not really listen to its environment and does not necessarily adapt to it, or very little. It does not change its qualities according to a relationship of proxemics, and therefore has no semantic relationship to space. One could say that it has a connection without relationship, since, in clinical psychophysiology, for there to be a relationship

---

[7]See the research projects led by Antonio Camurri (Università di Genova, Italy). Starting from the mapping of «emotional indexes» embodied in the body movement, Camurri has developped since 1999 the software EyesWeb and EyesWeb XMI Gesture and Social Processing Software Library. See Camurri [7, 8], et A. Camurri, B. Mazzarino, G. Volpe (2004) Analysis of Expressive Gesture: The EyesWeb Expressive Gesture Processing Library, in A. Camurri, G. Volpe (eds.), Gesture-based Communication in Human-Computer Interaction, LNAI 2915, pp. 460–467, Springer Verlag.

[8]The "gestosphere" or "sphere of the gesture" is proposed by Hubert Godard as the ensemble of connections (physical, perceptive, affective, and symbolic) that relate people to their environment. In contrast to Laban, Godard insists on the way the "dynamosphere" is psychologically and symbolically experienced by a person. We could therefore advance that the Godardian gestosphere is based on the dynamosphere, but goes beyond it from the viewpoint of meaning, since it immediately takes into account the psychological experience of the person. Cf. L. Louppe, *Poétique de la danse contemporaine*, Brussels, Contredanse, 1997, pp. 68–70.

(and not only a connection), consideration and hearing of the other in a shared space is necessary. A robot without "sensory organs" (i.e., sensors) is trapped in a geometric and functional fundamentalism in which listening to the other is subservient to its functions, its program. And it is not capable of adapting temporality or geometry to its task. One could speak of a perverse relationship to the other, the perversion being the non-recognition of the other as an other.

For a humanization of behavior, to move from the dynamosphere to the gestosphere, the robot must be endowed with two things in addition to the ability to move: "sensory organs" and an axiology, meaning a system of perceptual categorization that allows to assign meaning, to establish priorities, to know how to focus on certain important aspects of the environment and to know how to ignore others, according to circumstances and setting. The organs must sense the world, but it is the attribution of meaning to the perceived elements that allows for the definition of thresholds for the triggering and the accomplishment of actions. That attribution of meaning is certainly what determines emotion vis-à-vis the world. Emotion seeps through body, through pre-movement and the dynamics that color and provide "timbre" to the movement, building the layers of its harmonics, transforming it into a gesture, which finally gives it its meaning. The meaning given to the environment gives shape to our senses which in turn give direction to our movement (sens) by transforming it into behavior.

In summary, a three-tiered structure is once again taking shape:

1. the kinesphere, based on the geometry and motion that define the movement
2. the dynamosphere, based on the energy accents that define the gesture
3. the gestosphere, which is to say the relationship of a gesture to the context that defines the behavior

Consequently, can we use that general model to describe gesture in order to move closer to an implementable system? And can that model be compatible with the one elaborated in psychology of action?

3 Towards a System of Notation and Implementation of Behavior

We are attempting to identify the main features of what could constitute a system of notation and implementation of behaviors for non-anthropomorphic, non-zoomorphic robotized objects. Our goal is to go beyond a simple description of movement in the form of linear scores evoking motions in space by integrating the qualitative, relational, and behavioral aspects previously mentioned. We have retained two approaches to evoking the relationships among motion in space, semantics of movement and contextualization of behavior. The first is a product of the psychology of the perception of action, and allows us to better understand the way the psychological properties spontaneously attributed to an object endowed with movement are organized. The second seeks to define what, beyond the geometry of movement, constitutes a gesture in its relationship to the postural, dynamic, and contextual organization of the body. As for the notion of an agogics of movement, it puts forward the texture of movement in its idiosyncratic dimension.

3.1 Notating Movement to Attempt to Implement Behaviors

Notation, be it choreographic or even musical, makes it possible to formalize and transcribe in a conventional form (established by a code) the components of a work and, more specifically here, a series of notes or actions to be accomplished in order to produce or to instantiate the work in question. Consequently, that notation is shareable with a community that has knowledge of the conventions and can thus perform the work in accordance with that notation. The notation makes it possible to store, distribute, and perform the work just as it makes possible the creation of works without carrying them out, remaining in an "ideal" phase. Those works are then allographic in nature [19, 20], principally characterized by a two-stage operation: the time of notation, then that of instantiation, which then leads to the realization of the work. That dual temporality calls to mind another, so common today: that of programming and its running by a computer. What are the fundamental differences between these systems of executable notation? How and why could we bring these two worlds that share the same dual-time system closer together? And to what extent is it possible?

If computer programming is an increasingly common practice, it is a product of an abstraction that does not always make it possible to comfortably reconstruct or work out, even mentally, what running it will produce. Many programming environments therefore function in layers, from lowest, very abstract and detailed, to much more concrete, even figurative, levels that make it possible to have a global vision. The latter include auteur-oriented software such as graphic programming environments that work through patches (such as Max MSP, Pure Data, Isadora, etc.) or with command boxes such as Scratch (from the MIT Media Lab) followed by Snap (from Berkeley) or the Lego Mindstorm software, or even Blocky. All this software are also based on a system of graphics objects that can be snapped together like bricks or pieces of a puzzle.

More specifically within the scope of our research—though dedicated to Aldebaran robots (NAO, Pepper, and Roméo, all three of which are eminently anthropomorphic)—the software environment presented by Aldebaran (around the NAOqi OS) is very interesting, particularly Chorégraphe, a graphics programming authoring software itself using a system of "blocks" that can be connected amongst themselves to create a "flow diagram" and behaviors (or that can use a library of already available behaviors), with the possibility of publishing (via Curve) movements via a TimeLine in accordance with logic stemming from animation, or by publishing those movements through programming.

Bringing together choreographic notation and graphic programming environments would make it possible to manipulate a system of signs relating to the concrete physical reality of bodies while making the system executable by a computerized machine requiring an abstract logical system, a program. Such a notation/programming environment for movements, even behaviors, would make it possible to compose scores for objects at the meeting point of the two worlds in question (choreography and computers), for robots, especially in human form (humanoid or android).

We clearly see the value of such executable scores, which could be produced by authors equally to store or transmit or to have a humanoid robot produce a series of actions by using a symbolic graphic system, though also by analogy to the human body, like the main choreographic notations (Laban and Benesh).

However, that perspective presents two problems, one general and the other specific to our approach. The first relates to the fact that the computer and robotic paradigm is primarily interactive before it is linear. To have a robot perform a linear score of movements is of very limited interest when one considers it from the perspective of robotics or more broadly from interaction design. How can openings, variables, or a conditional principle be integrated into the notation system? That issue is absolutely central, including specifically for our approach, because that is also a question when it comes to passing from movement to behavior, something that is primarily relational and contextual.

The second problem is more directly related to our approach: what about when one is interested in non-anthropomorphic—even non-zoomorphic—objects? Even objects with no limbs—abstract or utilitarian objects, for instance? How then to take from these notation systems the dimensions that make possible the notation of movement not according to bodily positions, but according to qualities transposable to other forms and interpretable as concerning psychological, behavioral characteristics?

Another fundamental difference between the performance of a score by a human being and the running of a program by a computer obviously resides in the quality: variable and dependant on the humanity of the performer and (usually) invariable for a machine. As we have seen, in performance there is a dimension that goes beyond respect for the score, the notation, and which is perceptible precisely in the manner, the style, the energy, the agogics, etc. with which the performers take possession of the score, producing an instantiation as respectful of the work as it is singular. That performance, if we follow its trail, makes it possible to describe the author, to feel certain traits of their personality, their character, their way of functioning, their behavior. Where the performer, a human, makes gestures, the robot makes movements. How can we appeal to the intrinsic quality of those movements and to their relational and contextual qualities, such that those mechanical movements can have similarities with gestures, or even fall within the province of behavior, including for non-anthropomorphic, limbless robots?

To do that, we pose the hypothesis of this twofold approach: one part (psychology of action) starting from human perception of movements and the resulting attribution of behaviors (from animacy to social agency), and the other (gestural analysis) starting from movement and going back toward its context (from the kinesphere to the gestosphere). The question then is how to bring those two approaches together in a model allowing, in the end, for an implementable formalization, while factoring in two imperatives: that our system be conditional (open to contextual conditions) and applicable to non-anthropomorphic and non-zoomorphic objects.

3.2 Shape the Movement with a Set of Constraints

As we have seen in this article, the shift from movement to behavior—that is from a purely geometrical description of movements to a description integrating psychological attributes as well as a spatial and temporal context—necessitates several steps: an account of body posture and its relation to terrestrial gravity (pre-movement); a description of the dynamic movement organization, calling forth the notions of effort and energy, specifically in the context of the dynamosphere and agogics; and the representation by an observer of certain properties of the scene—physical constraints and psychological properties. To summarize this approach, we could define three dimensions that may organize the notation system we are looking for:

– **Movement**: This is the description of movements within a kinematic model, that is without referring to causes of motion, in terms of relative positions within a system of reference and differential properties (speed, acceleration). In the domain of dance notation, this dimension corresponds to the **kinesphere**, as described by Laban;
– **Gesture**: With respect to a notation based on geometric properties, a qualitative notation introduces certain attributes that do not depend on the linear progress of a score, but leaves room for certain fluctuations, the specific texture of a performance (agogics). This corresponds to the dynamosphere, based on energetic emphases;
– **Behavior**: The behavioral dimension in notation integrates different possible junctions of a movement sequence, its different evolutions depending on contextual variations. This dimension corresponds to the gestophere, the sum of projective and symbolic relationships to the environment, before and during the movement.

Those dimensions do not strictly correspond to the psychological organization we elaborated in Chap. 2: animacy, agency, and social agency. In fact, the three dimensions, movement, gesture, and behavior, are present at each of those levels. However, they do not bear the same weight. For instance, the movement dimension is more critical at the animacy level, inasmuch as the objects described at that level do not yet possess intentional properties that would allow the connection of their displacements to non-physical attributes. The behavior dimension, in its symbolic aspects, is more critical at the social agency level, insofar as the social interaction abilities may enable communication behaviors to take place.

To progress in the direction of a system of notation able to integrate these different aspects, the movement dimensions should be considered in relation to the psychological mechanisms engaged by an observer when perceiving the movement of an object. Some quantifiable motion features (direction, speed, or a change of volume of an object) should be linked to the scene features that an observer mentally represents and that may constrain the interpretation of the motion (for instance whether the observer is aware of certain forces that organize the motion, or of certain computational characteristics of the robot that determine the way it behaves). To move forward in that direction, we could conceive of a system of constraints that would determine the actualization of a movement in a robotic artifact and the interpretation given by an observer.

As we have seen, authoring environments rely on relatively abstract graphic interfaces and allow to program with patches, building blocs or following a timeline. Those different visual paradigms have already proved themselves useful, but we are looking for an even more intuitive solution that could go a step further in the representation of a behavior. We are just contemplating hypotheses that need to be developed and confirmed, but right away we want to consider some possible orientations. We assume that behavior, as expressed through evocative movements, is the result of different forces and constraints that we could formalize as a sort of landscape expressing the arrangement of these parameters. If this approach follows the line of Laban's «spheres» and graphic programming environments, the goal here is to integrate behavioral properties that could elicit specific psychological attributions, through a graphic interface with a high-level of abstraction. To configure a behavior as a landscape of forces and constraints, the idea is to create a dynamic and contextual cartography that takes into account the internal states of the objects as well as the context in which it evolves. Take for instance an object that moves on its own but tends to be scared by people around him. Assuming that this object is going to slow down in the vicinity of people, this inhibition could be transcribed in a landscape of force and constraints as a slope to climb, with a slope angle varying as a function of the degree of fear we want to impart to the object's behavior. The degree of inhibition is therefore configured as a constraint on movement, that is a resistance opposed to the initial object's velocity. Depending on the context and the type of object we want to animate, depending also on the degree to which this robotic object can be influenced by its environment, these constraints could also be external. This system of configuration by constraints amounts to a constellation of forces, both internal and external, from which a movement can result. This system could obviously be dynamic and adaptable, with different parameters to configure.

To progress on that idea of internalized landscape as a way to promote an abstract graphic representation of behavioral properties, we propose a hierarchy of constraints. Constraints at a superior level in a hierarchy would organize the way those at inferior levels act on the movement parameters (for instance, the intention to walk in a particular direction, a behavioral constraint, determines the postural

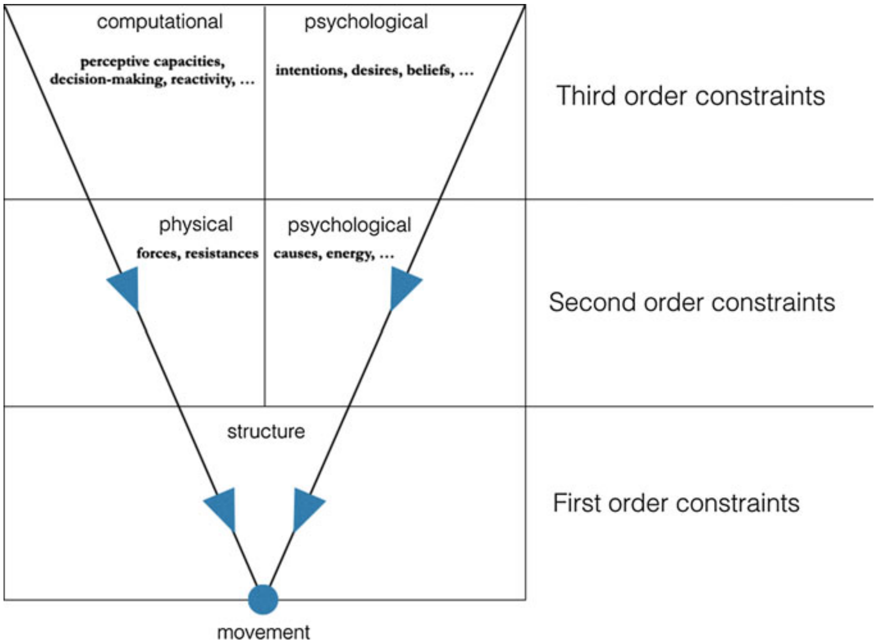


Fig. 5 Three orders of constraint to elaborate a robotic artifact's movement

adjustment and the dynamic realization of the movement, as a function of the ground). In that way, we can represent the process of conceiving and implementing expressive movements in a robotic artifact as the determination of variables at three levels of constraints (Fig. 5):

– **first-order constraints**: These constraints correspond to which movements are possible given the physical structure of the robot, the part it is composed of, and their degree of freedom.
– **second-order constraints**: These constraints correspond to the physical forces and resistances that may intervene during the realization of a robot's behavior and shape its execution. This type of constraint is at an intermediate level between the physical implementation of a robot and the behavioral properties we want it to display. They should make it possible to describe the movement's pace as a function of temporary energetic asymmetries that are resolved as the behavior unfolds. These constraints can be conceived as virtual forces and resistances inside an internalized landscape, meaning they need not be part of the actual physical world the robot is embedded into. Rather, they are meant to suggest to a human observer a particular dynamic organization, by interacting with her own intuitive physics and capacities of motion interpretation, in relation to contextual cues in the scene. For instance, an additional resistance applied to the robot's motion would produce an impression of sluggishness and overall difficulty to evolve smoothly. Other example: a constraint specifying an

external force resulting in a sudden acceleration and deviation from an initial trajectory, with a constant direction, may suggest, depending on the possibility to relate the deviation to an external landmark in the scene, that the robot is running away from something or that it is eagerly attempting to reach something.

– **third-order constraints**: These constraints correspond to the behavioral capacities that we want to be attributed to a robot by an external observer. These capacities need not be part of the actual computational endowment of the robot, but are to be suggested through consistent motion patterns and the interaction of the robot with its environment. The third-order constraints may correspond to at least three categories of cognitive abilities: perceptive abilites (to what extent the robot is aware of its surrounding), reactivity (to what extent the robot is prompt to change its behavior in reaction to external events), and decision-making (to what extent the robot's behavior is goal-oriented and flexible). Those abilities, expressed by constraints on the robots' movements, would be likely to receive a translation in terms of naïve psychology, that is according to intentions, desires and beliefs, as well as other psychological traits. For instance, constraints that specify the degree to which the robot's behavior is impacted by external events would determine an impression of awareness, curiosity, or indifference. Constraints related to the consistency of its behavior over time would influence the possibility for an observer to identify goals and determine how persistent the robot appears.

4 Conclusion: A Dual Interpretation to Analyze and Implement Behaviors

To analyze, conceive, and implement behaviors in objects that do not look like living beings, we have elaborated on two perspectives on the interpretation of movement. The first perspective corresponds to the psychological inferences spontaneously drawn when observing a movement in space. The second perspective is based on dance notation and the notion of emphasis in a musical phrase. We looked for a way to go beyond existing notation systems based on the human body [26], a system that could extract expressive movements that apply independently of the physical structure of the object, and contribute to a system of notation for non-anthropomorphic robotized objects. We delineated a solution that takes into account the human ability to interpret motion characteristics in terms of psychological traits, and we proposed a progression from movement to gesture as a way to move towards the description of genuine behavioral qualities and expressive attitudes and emotions. From these two perspectives, we proposed two typologies, both organized around three levels: animacy, agency, and social agency on the one hand; kinesphere, dynamosphere, and the gestosphere in the other. If these two typologies do not correspond directly, they can be joined based on a series of constraints at different levels of abstraction. These constraints, conceived as an arrangement of forces and constraints in a sort of internalized landscape, could contribute to the representation of behavior inside a graphic programming environment that takes movement notation beyond a mere linear score.

Because we propose a model that is based on constraints rather than on direct specifications, on conditions rather than on descriptions, it may be important to integrate a system of notation that would guide the implementation of movement in a robotic artifact. This system of notation could be composed of a library of movements, called forth as a function of a set of constraints defined upstream. This library should be composed of two elements: a system of movement notation with descriptions of the relative motions both in space and time, and a "qualitative" score that would, on the model of what is done in music notation, give specific performance instructions. This qualitative score could borrow from Labanotation, especially the Effort Shape notation, and/or from the agogic principles of notation in music.

References

1. J.A. Baird, D.A. Baldwin, Making sense of human behavior: action parsing and intentional inference, in *Intentions and Intentionality: Foundations of Social Cognition*, eds. by B. Malle, L.J. Moses, D. Baldwin (MIT Press, 2001), pp. 193–206
2. S. Baron-Cohen, How to build a baby that can read minds: cognitive mechanisms in mindreading. Cah. de Psychol. Cogn. Current Psychol. Cogn. **13**, 1–40 (1994)
3. C. Bartneck, T. Kanda, O. Mubin, A. Al Mahmud, Does the design of a robot influence its animacy and perceived intelligence? Int. J. Social Robot. **1**(2), 195–204 (2009)
4. S. Bianchini, R. Bourganel, E. Quinz, F. Levillain, E. Zibetti, (Mis)behavioral objects: empowerment of users vs empowerment of objects, in *Empowering Users through Design. Interdisciplinary Studies and Combined Approaches for Technological Products and Services*, ed. by D. Bihanic (Springer, Berlin, 2015)
5. C. Breazeal, A. Brooks, M. Hancher, J. Strickon, C. Kidd, J. McBean, D. Stiehl, Public anemone: an organic robot creature, in *ACM SIGGRAPH 2002 Conference Abstracts and Applications. SIGGRAPH '02* (San Antonio, Texas, 21–26 July 2002), pp. 76–76
6. C. Breazeal, A. Brooks, J. Gray, M. Hancher, J. McBean, D. Stiehl, J. Strickon, Interactive robot theatre. Commun. ACM **46**(7), 76–85 (2003)
7. A. Camurri, S. Hashimoto, M. Ricchetti, R. Trocca, K. Suzuki, G. Volpe, EyesWeb—toward gesture and affect recognition in interactive dance and music systems. Comput. Music J. **24**(1), 57–69 (2000)
8. A. Camurri, EyesWeb, in *La scena digitale, nuovi media per la danza*, eds. by A. Menicacci, E. Quinz (Venezia Marsilio, 2000)
9. A. Camurri, G. Volpe (eds.), G*esture-Based Communication in Human-Computer Interaction*. LNAI 2915. (Springer, Berlin, 2004)
10. G. Castellano, A. Camurri, M. Mortillaro, K. Scherer, G. Volpe, Espressive gesture and music: analysis of emotional behaviour in music performance. Music Percep. **25**(6), 103–119 (2008)
11. M. Chion, *Guide des objets sonores* (Pierre Schaeffer et la recherche musicale. Paris Buchet/Chastel, INA, 1983)

12. G. Csibra, G. Gergely, The teleological origins of mentalistic action explanations: A developmental hypothesis. Dev. Sci. **1**(2), 255–259 (1998)
13. N. Epley, A. Waytz, J.T. Cacioppo, On seeing human: a three-factor theory of anthropomorphism. Psychol. Rev. **114**(4), 864–886 (2007)
14. J. Forlizzi, C. DiSalvo, Service robots in the domestic environment: a study of the roomba vacuum in the home, in *Proceedings of the 1st ACM SIGCHI/SIGART Conference on Human-Robot Interaction, HRI'06* (ACM, New York, 2–4 Mar 2006), pp. 258–256
15. H. Fukuda, K. Ueda, Interaction with a moving object affects one's perception of its animacy. Int. J. Social Robot. **2**(2), 187–193 (2010)
16. T. Gao, G.E. Newman, B.J. Scholl, The psychophysics of chasing: a case study in the perception of animacy. Cogn. Psychol. **59**(2), 154–179 (2009)
17. S.A Gelman, G.M. Gottfried, Children's causal explanations of animate and inanimate motion. Child Dev. **67**, 1970–1987 (1996)
18. H. Godard, Le geste et sa perception, in *La danse au XX^{ème} siècle*, eds. by I. Ginot, M. Michel (Larousse, Paris, 2008), p. 236
19. N. Goodman, *Language of the Arts* (Indianapolis, New York, 1968)
20. N. Goodman, in *Of Mind and Other Matters*. (Harvard U.P, Cambridge, 1984)
21. J.K. Hamlin, K. Wynn, P. Bloom, Social evaluation by preverbal infants. Nature **450**, 557–560 (2007)
22. F. Heider, *The psychology of interpersonal relations* (Wiley, New York, 1958)
23. B. Hendriks, B. Meerbeek, S. Boess, S. Pauws, M. Sonneveld, Robot vacuum cleaner personality and behavior. Int. J. Social Robot. **3**(2), 187–195 (2011)
24. M.L. Hoffman, *Empathy and Moral Development: Implications for Caring and Justice* (Cambridge University Press, Cambridge, 2001)
25. G. Hoffman, W. Ju, Designing robots with movement in mind. J. Human Robot Interac. **3**(1), 89–122 (2014)
26. M. Karg, A.A. Samadani, R. Gorbet, K. Kuhnlenz, J. Hoey, D. Kulic, Body movements for affective expression: a survey of automatic recognition and generation. Affect. Comput. IEEE Trans. **4**(4), 341–359 (2013)
27. R. Laban, Choreutics, pp. 27–36. Dance Books Limited, London (2001) (1966 orig. ed.)
28. R. Laban, F.C. Lawrence, *Effort Shape* (Mc Donald and Evans, London, 1947)
29. P. Lalitte, Conditions de possibilités d'une rhétorique formelle perçue. Intellectica **1–2**, 103–114 (2008)
30. M. Leman, *Embodied Music Cognition and Mediation Technology* (The MIT Press, Cambridge Massachusetts, 2007)
31. F. Levillain, S. Lefort, E. Zibetti, Moving on its own: how do audiences interact with an autonomous moving artwork? in *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems*, *Proceeding of CHI 2015* (ACM, 2015), pp. 695–702. http://dl.acm.org/citation.cfm?id=2702973
32. B.F. Malle, Attribution theories: how people make sense of behavior, in *Theories in Social Psychology*, ed. by D. Chadee (2011), pp. 72–95
33. A. Menicacci, E. Quinz, *La scena digitale, nuovi media per la danza* (Marsilio, Venezia, 2000)
34. A. Menicacci, E. Quinz, *Etendre la perception? Biofeedback et transfert intermodaux en danse*. in Scientifiquement danse, Nouvelles de danse 53, Bruxelles (2005), pp. 76–96
35. A. Michotte, *La Perception de la Causalité* (Publications Universitaires de Louvain, 1954)
36. D. Newtson, Attribution and the unit of perception of ongoing behavior. J. Pers. Soc. Psychol. **28**, 28–38 (1973)
37. M. Noma, N. Saiwaki, S. Itakura, H. Ishiguro, Composition and evaluation of the humanlike motions of an android, in *2006 6th IEEE-RAS International Conference on Humanoid Robots* (IEEE, 2006), pp. 163–168
38. A. Paiva, J. Dias, D. Sobral, R. Aylett, P. Sobreperez, S. Woods, C. Zoll, L.E. Hall, Caring for agents and agents that care: building empathic relations with synthetic agents, in *Proceedings of the Third International Joint Conference on Autonomous Agents and Multiagent Systems*, vol. 1 IEEE Computer Society, 2004), pp. 194–201

39. A. Pereira, I. Leite, S. Mascarenhas, C. Martinho, A. Paiva, Using empathy to improve human-robot relationships, in *Human-Robot Personal Relationships* (Springer, Berlin, 2011), pp. 130–138
40. R.W. Picard, Affective computing: challenges. Int. J. Hum Comput Stud. **59**(1), 55–64 (2003)
41. D. Premack, The infant's theory of self-propelled objects. Cognition **36**(1), 1–16 (1990)
42. S.D. Preston, F. De Waal, Empathy: its ultimate and proximate bases. Behav. Brain Sci. **25** (01), 1–20 (2002)
43. J. Saez-Pons, H. Lehmann, D.S. Syrdal, K. Dautenhahn, Development of the sociability of non-anthropomorphic robot home companions, in *2014 Joint IEEE International Conferences on Development and Learning and Epigenetic Robotics (ICDL-Epirob)* (IEEE, 2014), pp. 111–116
44. B.J. Scholl, P.D. Tremoulet, Perceptual causality and animacy. Trends Cogn. Sci. **4**, 299–309 (2000)
45. P. Shaeffer, *À la recherche d'une musique concrète* (Seuil, Paris, 1952)
46. P. Shaeffer, *Traité des objets musicaux* (Seuil, Paris, 1966)
47. P. Tremoulet, J. Feldman, Perception of animacy from the motion of a single object. Perception **29**(8), 943–952 (2000)
48. P. Tremoulet, J. Feldman, The influence of spatial context and the role of intentionality in the interpretation of animacy from motion. Percept. Psychophys. **68**(6), 1047 (2006)
49. F. Voisin, LOL: Un environnement expérimental de composition chorégraphique, Paris, Ec/cart, vol. 2, Eric Sadin ed., Dif'Pop (2000)
50. J.E. Young, J. Sung, A. Voida, E. Sharlin, T. Igarashi, H.I. Christensen, R.E. Grinter, Evaluating Human-Robot Interaction. Int. J. Social Robot. **3**(1), 53–67 (2010)
51. J.M. Zacks, Using movement and intentions to understand simple events. Cogn. Sci. **28**(6), 979–1008 (2004)

Laban Movement Analysis and Affective Movement Generation for Robots and Other Near-Living Creatures

Sarah Jane Burton, Ali-Akbar Samadani, Rob Gorbet and Dana Kulić

Abstract This manuscript describes an approach, based on Laban Movement Analysis, to generate compact and informative representations of movement to facilitate affective movement recognition and generation for robots and other artificial embodiments. We hypothesize that Laban Movement Analysis, which is a comprehensive and systematic approach for describing movement, is an excellent candidate for deriving a low-dimensional representation of movement which facilitates affective motion modeling. First, we review the dimensions of Laban Movement Analysis most relevant for capturing movement expressivity and propose an approach to compute an estimate of the Shape and Effort components of Laban Movement Analysis using data obtained from motion capture. Within a motion capture environment, a professional actor reproduced prescribed motions, imbuing them with different emotions. The proposed approach was compared with a Laban coding by a certified movement analyst (CMA). The results show a strong correlation between results from the automatic Laban quantification and the CMA-generated Laban quantification of the movements. Based on these results, we describe an approach for the automatic generation of affective movements, by

S.J. Burton (✉)
Sheridan College, Oakville, Canada
e-mail: bsjdanceresearch@gmail.com

A.-A. Samadani
Institute of Biomaterials and Biomedical Engineering, Bloorview Research Institute,
University of Toronto, Toronto, Canada
e-mail: ali.samadani@utoronto.ca

R. Gorbet
Department of Knowledge Integration, University of Waterloo, Waterloo, Canada
e-mail: rbgorbet@uwaterloo.ca

D. Kulić
Department of Electrical and Computer Engineering, University of Waterloo,
Waterloo, Canada
e-mail: dana.kulic@uwaterloo.ca

J.-P. Laumond and N. Abe (eds.), *Dance Notations and Robot Motion*,
Springer Tracts in Advanced Robotics 111,
DOI 10.1007/978-3-319-25739-6_2

adapting pre-defined motion paths to overlay affective content. The proposed framework is validated through cross-validation and perceptual user studies. The proposed approach has great potential for application in fields including robotics, interactive art, animation and dance/acting training.

1 Motivation

Every day we interpret others' expressions by observing their body language. Suppose only the arm and hand are visible; there is no head, face, or torso available for any revealing clues. Now imagine that the arm and hand have been transformed into a sculptured frond. It consists of many joints with bits of flexible plastic feathery "muscles" attached. It moves in hand-like waves. Does it affect you? Is it possible to translate emotional expression through the body, via algorithms, into a moving sculpture "with feelings"? This sculpture and associated questions illustrate our motivation. It is one of the immersive, responsive sculptures in a series entitled: *Hylozoic Ground*, by Philip Beesley and Rob Gorbet. In 2010, *Hylozoic Ground* was Canada's selection at the Venice Biennale of Architecture (Fig. 1).

Hylozoism is an ancient belief that all matter is sentient. The sculpture's movements are affected by the proximity of the viewer, who becomes both audience and participant. The sculptural installations consist of a large number (from dozens to hundreds) of sensors and actuators; from the robotics perspective, the sculpture can be considered as a robot with a large number of degrees of freedom. However, unlike many robots, it is not anthropomorphic, but rather, emulates non-human natural forms, akin to a forest canopy. The sensors detect the presence and proximity of visitors and generate movements and other activation in response.

Fig. 1 Illustration of the Hylozoic series sculpture, frond mechanisms in the foreground. Photograph by Philip Beesley, reprinted with permission

Currently, the motion generation strategy is very simple, consisting of open loop control with a random component. However, visitors do not perceive these movements to be random, rather they perceive that the movement has an emotional (affective) content, e.g. “the sculpture was happy to see me”, “it was sad” [1]. This was initially an unexpected element of the interaction. We wondered if this communication path could be used more extensively by the installation designer to choreograph a movement with an intended affective content. Since people move in response to the sculpture, would it be possible to observe their movements and interact affectively through movement? This capability may be valuable beyond artistic installations, in applications such as human-computer interaction and human-robot interaction. Our goal then was to develop a way to translate between the *movement language* of humans and the potential *movement language* of the sculpture.

As a path to reach our goal, our research focused on affective human gestures of the arm and hand, the parts of the body that would be most similar to the sculptural fronds. Variations in expressions of emotion are dependent on each individual’s *body-mind*, a word coined by Bonnie Bainbridge Cohen [2] used to reinforce the fact that the body and mind are linked together within a person. Each individual has a unique personal history that influences their movements, as does their physical construction and ability. The challenge of this study, this partnership between the science of robotics and the art of expressive movement, was to attempt to discover and distil the essence of affective movement. The engineers looked to the dance/theatre performance world, where choreographed movements are specific and repeatable with believable affective qualities, for a language to analyze and describe movement. The photographs in Fig. 2 illustrate expressive arm/hand movements choreographed to create affective responses in a theatre audience.

Fig. 2 *Left photo* Illustrates an expansive, all-encompassing joy in dancer’s lightly held arms with controlled fingers miming repetitive quick, “talking” gestures; in contrast to the singer’s delicate precisely-focused appeal. *Right photo* Illustrates a light, sustained, fluid enclosing gesture of shy love in response to a gentle touch. (author’s personal collection)

Our approach aims to formalize and quantify the relationship between perceived movement qualities and measurable features of movements, to enable this relationship to be exploited for automated recognition and generation of affective movement. Another challenge of our research was to develop a common language and shared understanding of movement analysis between interdisciplinary research team members from the dance/choreography and engineering communities. In this monograph, we will describe our approach to motion understanding, based on insights from dance notation and computational approaches. In Sect. 2, we provide a brief overview of Laban Movement Analysis, focusing in particular on the Effort and Shape components, which are instrumental to the conveyance of affect. In Sects. 3 and 4 the design of the movement database and the data collection are described. Section 5 describes the analyst annotation procedure, which is used to generate ground truth for validating our proposed approach. Section 6 describes the proposed quantification approach and verification results. In Sect. 7, the use of the proposed approach for generating affective movements is illustrated. The Chapter ends with conclusions and directions for future work.

2 Laban Movement Analysis

I can't do much for you until you know how to see. -José de Creeft, sculptor

The study of Laban Movement Analysis (LMA) trains an observer to *see*, to become aware of, to attempt to ascertain the different aspects of movement. LMA promotes an understanding of movement from the inside out, as the mover, as well as from the outside in, as the observer. Rudolf von Laban (1879–1958) developed theories and systems of movement and notation. He wrote about the need to find a way to combine movement-thinking and word-thinking in order to understand the mental side of effort and action and re-integrate the two in a new form. When considering the expressive communication of the actor-dancer, Laban stressed that imitation does not "penetrate to the hidden recesses" of human inner effort. Laban searched for an authentic symbol of the inner vision in order for the performer to make effective affective contact with the audience, and felt that this could be achieved only if we have learned to think in terms of movement [3]. While attempting to capture movement in writing, he developed a system of basic principles and movement language that are encompassed in today's Laban Movement Analysis. Bloom argues "that LMA, by providing a vocabulary for articulating the detail of experiential phenomena, provides a valuable framework and a system of categories for bringing the interrelationships between body and psyche into greater focus." [4]. To enable automated movement analysis, a computational understanding of how affect is conveyed through movement was needed. Laban Movement Analysis was used to provide a language useful in the "translation of emotions to algorithms".

Laban Movement Analysis is divided into four overarching themes, both quantitative and qualitative. They comprise a blend of science and artistry. *Stability/Mobility* describes the natural interplay of components of the body that function to allow the full scope of human movement and balance to occur. *Exertion/Recuperation* speaks to the rhythms and phrasing of movements, that, similar to the rhythms of breath, may be said to create a "dance" between muscular tension and release. *Inner/Outer* addresses our connection from our needs and feelings within ourselves to our movement out in the world and the return flow of a response to our environment. *Function/Expression* differentiates between the aspects of movement that serve a need and the movement qualities that are expressive of affect. The latter two themes were of most interest to this project. There is some discussion amongst Certified Movement Analysts (CMAs) concerning the dichotomy between quantitative and qualitative analysis, assuming that concepts need to belong in one category or the other. The implication is that if something cannot be measured then it is qualitative and unprovable. The concepts in LMA are governed by principles, whether or not they are measurable, that make them "concrete, observable, experientially verifiable, repeatable and predictable." [5]. For this reason, we believe LMA is amenable to computational analysis and can be related to measurable features of movement.

Laban Movement Analysis employs a multilayered description of movement, focusing on the components: Body, Space, Effort and Shape. Body indicates the active body parts, and the sequence of their involvement in the movement; Space defines where in space the movement is happening, the directions, spatial patterns and range; Effort describes the inner attitude toward the use of energy; and Shape characterizes the bodily form, and its changes in space. If each of these aspects is understood in terms of its own integrity, one can begin to comprehend how each interacts and illuminates the others [5]. Irmgard Bartenieff (1890–1981), a colleague of Laban, advocates the use of Effort and Shape as a means to study movements from behavioural and expressive perspectives. Application of the concepts of quality, or "inner attitudes towards" movement, are used in the analysis of Effort [6]. Thus among Laban components, Effort and Shape are the most relevant for our study of affective movements.

The members of our research team, in order to communicate, needed to become familiar with each other's language, e.g., the terms "High Level-Low level" for the engineers referred to qualities of information but to the choreographer and actor, referred to placement in space. Symbols are international in a way that words are not. Laban Movement Analysis is the basis for both Labanotation and Laban Motif Notation. The choice for usage is generally based on the level of detail needed for the task at hand. Labanotation can include much detail for reproducing a movement sequence. For example, torso, shoulder, upper arm, lower arm, and separate finger gestures may be notated for precise reproduction purposes. Motif Notation is often used to capture the significant impressions, the similarities, the differences, and can lead readily to pattern recognition. In repetitive arm gestures, for example, it can be used to notate differences in various qualities or expressed efforts. Laban's terminology and symbols become meaningful with the consciously experienced

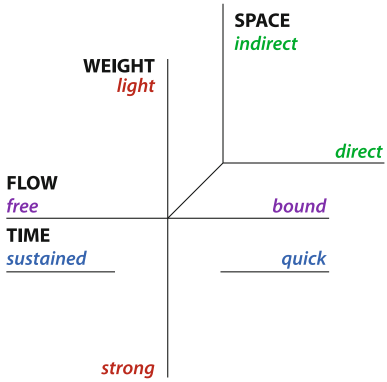


Fig. 3 The Laban Effort graph. The *short diagonal line* indicates Effort, and is part of every Effort symbol

embodiment of the specific movement quality. The symbols are derived from the basic Effort graph, illustrated in Fig. 3.

Table 1, adapted from Bartenieff [7] and Chi [8], illustrates different Effort qualities using simplified explanations of each of the Effort factors: Space ( ), Weight ( ), Time ( ), and Flow ( ), and their polarities or Effort elements: Space: Direct( )/Indirect( ); Weight: Strong( )/Light( ); Time: Sudden( )/Sustained ( ); and Flow: Bound ( )/Free ( ) [9]. A simple example of arm and hand gestures provides an illustration of each of the elements. Further examples can be found in Wile [10, p. 75] and Hackney [11, pp. 219–221].

Table 2 [11, 12] illustrates examples for Laban's Shape categories, known as Modes of Shape Change. Included are Shape Flow ( ) with two basic polarities: Growing ( )Shrinking ( ); Directional ( ) includes Arc-like Directional ( ) and Spoke-like Directional ( ) and Shaping or Carving ( ) which includes

Table 1 Laban Effort factors, adapted from Bartenieff [6] and Chi [7]

Effort factors	Elements	Example
Space ( ): attention to surroundings	Direct ( )	Pointing to a particular spot
	Indirect ( )	Waving away bugs
Weight ( ): sense of the impact of one's movement	Strong ( )	Punching
	Light ( )	Dabbing paint on canvas
Time ( ): sense of urgency	Sudden ( )	Swatting a fly
	Sustained ( )	Stroking a pet
Flow ( ): attitude toward bodily tension and control	Bound ( )	Carefully carrying a cup of hot liquid
	Free ( )	Waving wildly

Table 2 Laban Shape factors

Shape Factors	Elements	Example
Shape Flow ( ) is self-referential and defines readjustments of the body for internal physical comfort	Growing ( )	Self-to-self communication, stretching to yawn
	Shrinking ( )	Exhaling with a sigh
Directional ( ) is goal-oriented and defines the pathway to connect or bridge to a person, object, or location in space	Arc-like ( )	Swinging the arm forward to shake hands
	Spoke-like ( )	Pressing a button
Shaping/Carving is process-oriented and is the three dimensional "sculpting" of body oriented to creating or experiencing volume in interaction with the environment	Molding, contouring, or accommodating ( )	Cradling a baby

three-dimensional sculptural movements. Our focus in Shape was confined to Arc-like Directional and Spoke-like Directional [9]. Further discussion of Effort and Shape notation can be found in Dell [9] and Wile [10].

3 Designing the Movement Pathways

In designing the movement pathways, inspiration was taken from the types of movements similar to those of the fronds in the sculpture. The goal in designing the choreographed pathways was to choose several simple arm movements that were not already strongly weighted with affect, but were as neutral as possible. Michael Chekhov (1891–1955) is known for his development of what he called the psychological gesture for actors to use for character development. Lenard Petit in his book, The Michael Chekhov Handbook: for the actor [13], notes that the psychological gesture is an inner gesture, found with the physical body and archetypal in form, and that five archetypal gestures are used for training purposes: pushing, pulling, lifting, throwing and tearing. These are used as a means of realizing the "six statements of action" for an actor, which are "I want - I reject, I give - I take; I hold my ground - I yield" [13]. Each of the archetypal gestures can be done in each of the six directions: forward, backward, up, down, right and left. There is different information from each of these directions and there are an infinite number of qualities (adverbs) to work with [13]. This information closely allies with Laban's work. Genevieve Stebbins, in discussing Delsarte's system of expression in both theory and practice, explains that there are three things to be noted in order to fully understand the motions of the arm: (1) the articulations; (2) the attitudes and (3) the inflections [14]. Interestingly, under inflections are listed: declaration, negation, rejection, caress, affirmation, appellation, acceptation, attraction, and repulsion [14]. It is important to reinforce the fact that different factors such as culture, physique,

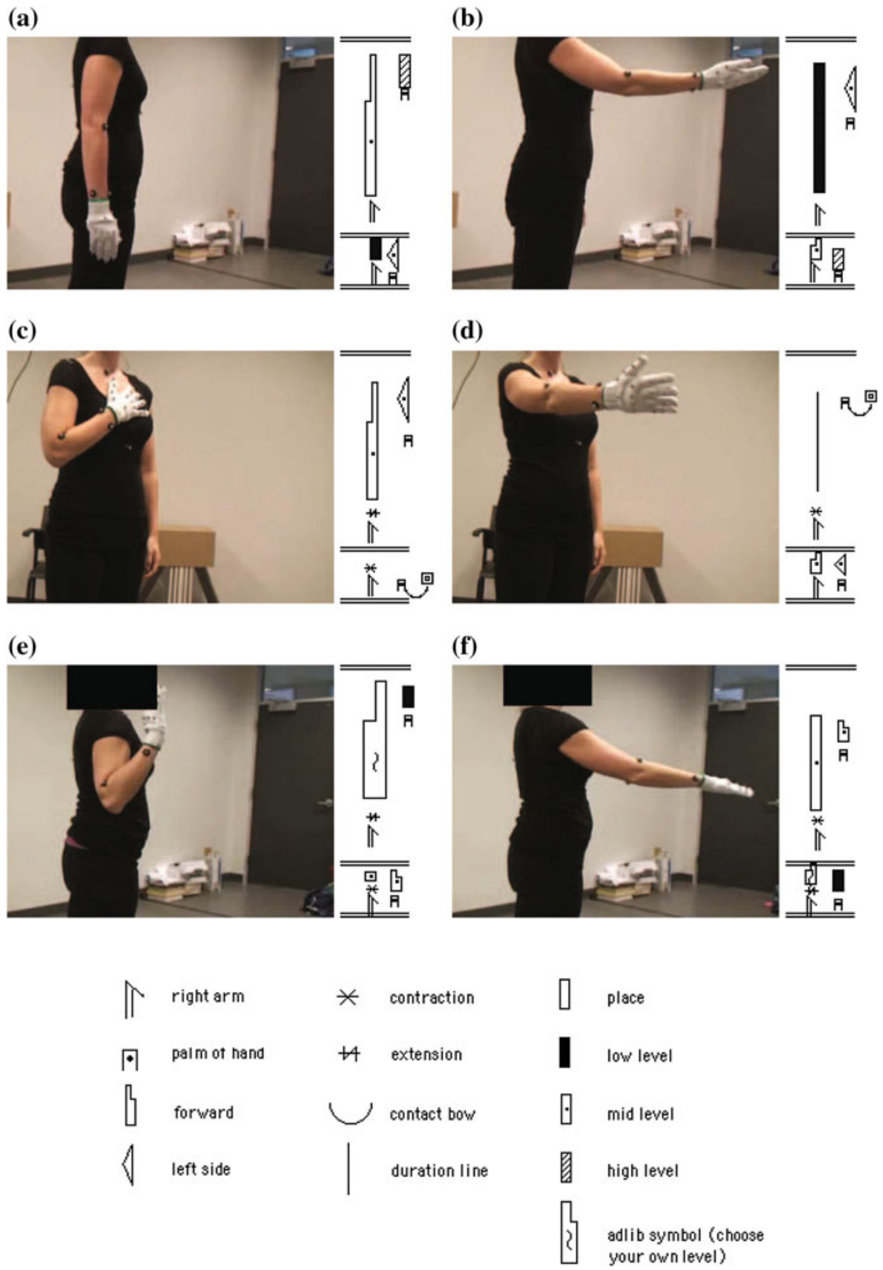


Fig. 4 The starting position and motif notation for each pathway. Motif notation and legend provided by Christine Heath. **a** pathway 1, **b** pathway 2, **c** pathway 3, **d** pathway 4, **e** pathway 5, **f** pathway 6, **g** motif notation legend

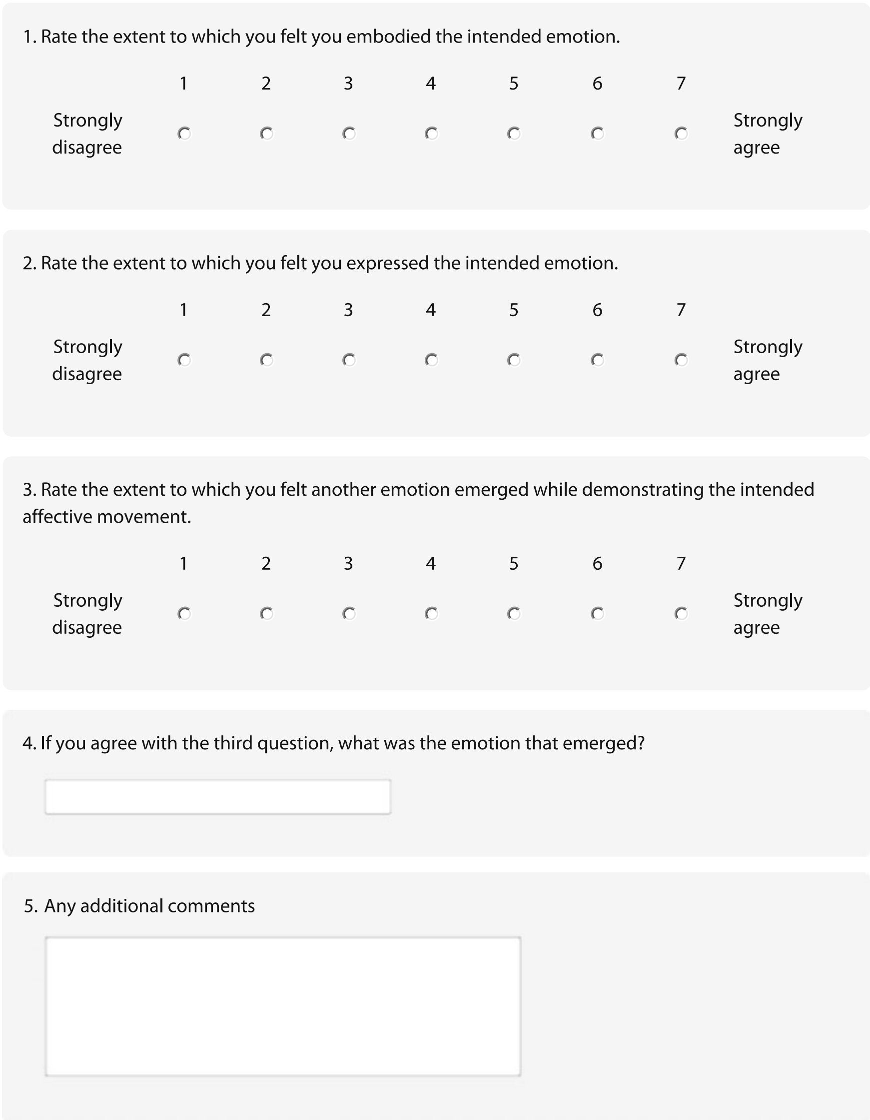
personal history, and specific environmental circumstances influence a quality of movement. North states that "[i]t is impossible to say either that a particular movement equals a special quality or that a particular quality equals one movement pattern plus a certain shape or space characteristic. Only generalizations can be made, because a movement assessment is made by the meticulous study of observed movement patterns of each individual." [15]. Physical experimentation augmented the study of the literature. Kent De Spain notes that using improvisation is a form of research. It is a means of delving into the complex natural system that is the human being. In a sense, movement improvisation is another way of thinking, one that produces ideas impossible to conceive in stillness [16].

Based on the study of gestures and accompanying experimentation, three simple pathways were chosen; each was also reversed, making a total of six pathways without strong affective associations. The more limited the prescribed pathway the higher the possibility of measuring subtle significant differences between the emotions. The actor's arm movements were to follow a given choreographed pathway in the first two of three sets. Direction along a pathway is usually significant, but this was not left free for the actor in the first and second sets. Palm facing and finger movements are also of emotional import, but variability was minimized, especially in the first set. A natural tendency to transfer expressive movements to other parts of the body, such as torso and shoulder, especially when the arm and hand movements were limited, would not be taken into account in this study due to the limitations of structure of the robotic frond.

Pathways: (1) The right arm starts down along the side, and moves up to forward mid-level, reaching, open palm facing up; (2) Similar motion as in (1) but in the reverse direction; (3) Starting with open palm on the chest, the right arm extends forward, ending with palm facing left, toward midline; (4) Similar motion to (3), but in the reverse direction; (5) Starting with the right arm bent with elbow down by the side, open palm facing forward in front of the right shoulder, the right arm extends forward at mid-level, open palm facing down, hand parallel to the floor and (6) Similar motion to (5), but in the reverse direction. (The latter two pathways were adapted by the actor to begin, or end, at a slight downward slant, between mid and low level.) The photos in Fig. 4 illustrate the beginning position and the Motif Notation for each pathway.

4 Motion Capture

For each of the six paths, the professional actor was asked to act each of Ekman's original Six Basic Emotions: anger, happiness, disgust, sadness, surprise and fear [17]. Prinz acknowledges that they have become the most widely accepted candidates for basic emotions, both psychologically and biologically [18]. With five tries for each emotion, we captured 180 movement sequences (6 paths, 6 emotions, 5 trials) for each of three data sets. For data set 1: The arm-hand follows the specified path, with an attempt at no extra wrist or finger movement other than just



1. Rate the extent to which you felt you embodied the intended emotion.

	1	2	3	4	5	6	7	
Strongly disagree	○	○	○	○	○	○	○	Strongly agree

2. Rate the extent to which you felt you expressed the intended emotion.

	1	2	3	4	5	6	7	
Strongly disagree	○	○	○	○	○	○	○	Strongly agree

3. Rate the extent to which you felt another emotion emerged while demonstrating the intended affective movement.

	1	2	3	4	5	6	7	
Strongly disagree	○	○	○	○	○	○	○	Strongly agree

4. If you agree with the third question, what was the emotion that emerged?

5. Any additional comments

Fig. 5 Questionnaire for the actor

an extension of the movement of the whole arm. For data set 2: the arm-hand follows the path, this time allowing movement of the wrist and fingers. For data set 3: the actor was allowed freedom of choice of arm and hand movements, including spatial pathway. This pilot study considered only data from data set 1.

Actors are rarely asked to "act" an emotion. Instead, actions and situations awaken emotions. The actor's ability to create a specific emotion in this setting was crucial to the success of this investigation. The actor relied on her rigorous training in the use of memory and imagination to freshly create and express the emotion aroused internally. In LMA, the word *intent* is used to describe part of the preparation stage of movement and "it is at this crucial point that the brain is formulating (even in a split second) the motor plan which will eventually be realized in action." [11]. As noted in Psychology of Dance, the more vivid, realistic and detailed the image is, the more the senses, thoughts and emotions are involved [19]. Directly following each group of five trials, the actor was asked for her personal response as to whether she felt she had or had not embodied the intended emotion, and if another emotion had emerged instead of the specified one, what was that other emotion. As Laban notes in the introduction to his book The Mastery of Movement, the variety of the human character derives from the multitude of possible attitudes toward the motion factors [3]. The training of the professional actor, the number of tries, and the questionnaire shown in Fig. 5, were attempts at providing high quality motion capture examples of the six emotions.

5 Analysis of the Motion Capture Videos

A coding sheet was devised for the Laban Certified Movement Analyst (CMA) to use while watching the video of each movement, shown in Fig. 6. Effort elements vary in intensity ranging from slight to marked [10]. A 5-point Likert scale was used: the "0" in the centre is checked when there appears to be no significant attention by the mover to that particular Effort. The "1 and 2" on either side of the "0", denote a noticeable increased degree of that Effort element. We focused on each of the Effort qualities: Time Effort: Sudden ( ) or Sustained ( ); Space Effort: Direct ( ) or Indirect ( ); Weight Effort: Strong ( ) or Light ( ), and Flow Effort: Bound ( ) or Free ( ). A 7-point Likert Scale was tried, but it was deemed too difficult, using video of only the arm, to translate this qualitative assessment into that much quantitative detail. There are three "levels" or opportunities for notating changes within a single movement; e.g. in Time Effort one may execute a Sudden impulse, to Timelessness or a steady time, to Sustainment or a slowing down. There is also a Comment Box for any explanatory notes deemed significant in the analysis, as shown in Fig. 6. For computing purposes, this scale was translated into a scale of 1 to 5. For the analysis of Shape, the focus was on Arc-Like ( ) or Spoke-like ( ) Directional.

We have omitted any discussion of Laban's States (a combination of two Efforts) and Drives (a combination of three Efforts) due to the fact that the computations were based on individual Effort elements and not their combinations.

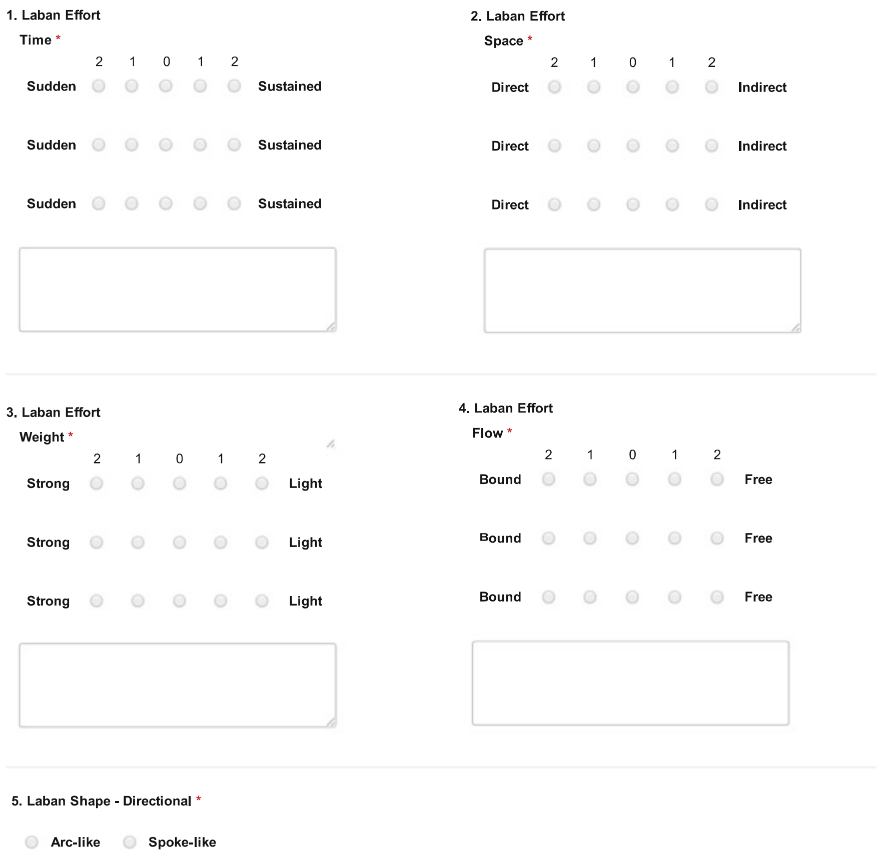


Fig. 6 Laban annotation questionnaire used by the CMA

6 Computational Laban Analysis

In addition to the longstanding research on movement analysis in the dance community, affective movement analysis has more recently also received significant attention in other domains. There is a large and active research effort on affective movement perception, recognition and generation in cognitive science, psychology, and affective computing [20]. Most closely related to our work, recently, two groups have proposed approaches for automated Laban Effort and Shape quantification. The first approach, proposed by Nakata and colleagues [21], developed a quantification approach for an aggregate set of body parts. The quantified components were used to generate dance movements, which were perceived by human observers to convey distinct affective expressions. This approach was later adopted by Hachimura et al. [22] for full-body movements, and applied to robot affective movement generation. The second approach, proposed by Kapadia et al. [23],

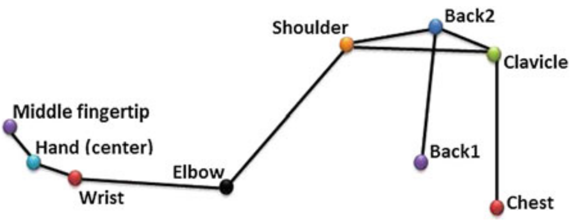


Fig. 7 Marker set used for the quantification

develops an approach for Laban Body, Effort and Shape quantification for individual body parts. Rather than using the approach for motion generation, Kapadia et al. use the quantification to optimize full-body motion indexing and retrieval from a motion database.

In both quantification approaches [21, 23], Laban descriptors are quantified as Boolean features,[1] therefore the quantification depends on the definition of suitable threshold. In our work [12], we propose a continuous quantification of LMA factors, using Nakata [21] and Kapadia [23] as the starting point. In the following, we label the approach using Nakata as the starting point as Q1, and the approach based on Kapadia as Q2. In addition, we propose quantification methods for dimensions which were not addressed by Q1 or Q2, namely a quantification for Shape: Shaping/Carving. Finally, we evaluate both proposed quantification approaches using the dataset described in Sect. 4, comparing the automated quantification with the labels generated by the expert analyst to verify the level of agreement.

Figure 7 illustrates the markers used for the quantification. These 9 markers are derived from the 30 markers used to collect data. The measured data consisted of 30 3D markers measured at 200 times per second, equaling nearly 20,000 data points for each second of video. The raw data is very high dimensional, providing a strong motivation to find a mapping between the raw data and a lower dimensional representation such as LMA, where movements can be more easily analyzed.

The first LMA factor quantified was Weight Effort, which describes the sense of force of one's movement, with the contrasting elements Strong and Light. Nakata et al. [21] proposed that Weight Effort be categorized based on a threshold on the kinetic energy for each body part. We adapt this approach in Q1 for a continuous valued quantification by estimating the maximum of the kinetic energy of the upper body parts, as illustrated in Fig. 8a. Kapadia et al. [23] proposed that Weight Effort be categorized based on a threshold of the deceleration of the different body part. We adapt this approach in Q2 so that the Weight Effort is

[1]A Boolean feature is one that can take on one of only two values, for example, True or False. In the case of LMA quantification, a Boolean feature means that each component is quantified as belonging to either one or the other of the extremum values, for example, for the component Weight Effort, each movement is classified as being either Strong or Light.

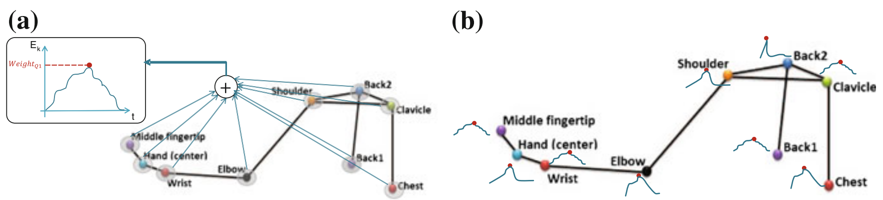


Fig. 8 Quantifications of weight effort. E_k is the kinetic energy

quantified as the maximum of the deceleration of the different body parts, as illustrated in Fig. 8b.

The second LMA factor quantified was Time Effort (_ ⸍_), which describes the sense of urgency, with the contrasting elements Sudden (⸍_) and Sustained (_ ⸍). Nakata et al. [21] propose that the quantification for this factor be based on the acceleration of different body parts. We adapt this approach in Q1 to propose that Time Effort be quantified as the peak of the sum of accelerations of the upper body parts weighted by their relative mass, as illustrated in Fig. 9a. Q2 proposed that the Time Effort be quantified as the net acceleration accumulated at the body parts, as illustrated in Fig. 9b.

The third LMA factor quantified was Space Effort (⅃—), which describes the attention to surroundings, with the contrasting elements Direct (⸝—) and Indirect (⅃). Nakata et al. [21] propose that the Space Effort be categorized by considering the relative direction between the torso and the face. This is implemented computationally by thresholding the inner product between the torso and face direction vectors. In Q1, this dimension is quantified by counting the number of peaks in the inner product of the tangents of the torso and wrist trajectories, as illustrated in Fig. 10a. This measure estimates how frequently the wrist trajectory changes direction relative to the torso. In Q2, Space Effort is quantified by computing the ratio of the total displacement and the net distance traveled by the body part, as illustrated in Fig. 10b.

The fourth LMA factor quantified was Flow Effort (⎯⦣), which describes the attitude towards bodily tension and control, with the contrasting elements Bound (⦣) and Free (⎯⸝). Flow Effort was not quantified in Q1; Q2 proposed a measure of

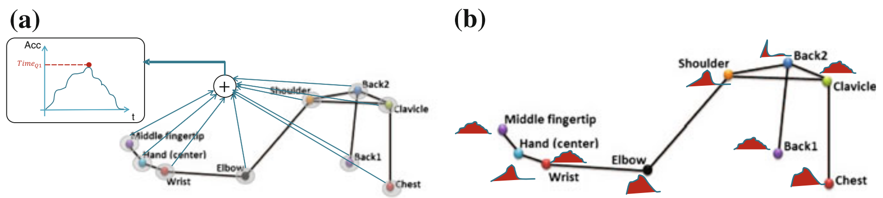


Fig. 9 Quantifications of time effort. *Acc* is the acceleration

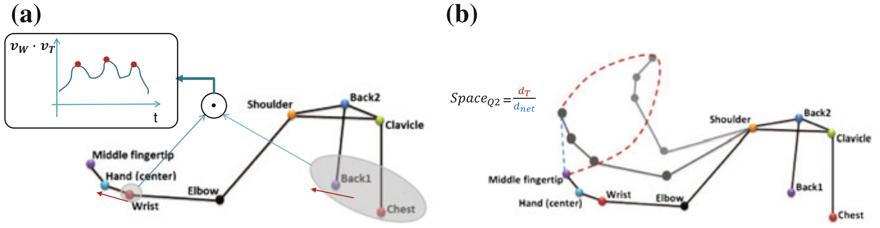


Fig. 10 Quantifications of space effort. v_W and v_T are the velocity of the wrist and torso. d_T is the total distance traveled, while d_{net} is the *straight line* (net) displacement from the starting to the ending pose

Flow Effort as the aggregated jerk over time at the considered body parts. Jerk is the rate of change of the acceleration of each body part.

The final LMA factor considered was Shape Directional ($\nrightarrow$), which defines the pathway to connect to or from the demonstrator to their goal in space, with the two categories of Arc-like ($\nrightarrow$) and Spoke-like ($\nrightarrow$). Neither Q1 nor Q2 proposed a quantification approach for this dimension. Here, we propose to quantify Shape Directional as the average curvature of the movement in a two dimensional plane within which the largest displacement occurs. The plane of movement is estimated using the top two components found via Principal Component Analysis.[2] The approach is illustrated in Fig. 11.

The proposed quantifications are evaluated by comparing the automated quantification values with the annotations provided by the CMA, as described in Sect. 5. A total of 44 hand and arm movements were annotated and the corresponding Effort and Shape factors quantified. As noted in Sect. 5, it was possible for the movement to contain multiple levels of each LMA component in a single movement, and the analyst had the opportunity to indicate this in her annotation by specifying multiple annotations. Movements with a variation in a single Effort factor, such as the previously mentioned example of Sudden impulse into even timing, into Sustained Time Effort, were not included. Only movements with a single annotation were considered for evaluation, to avoid the need to segment movements.

Two examples of the proposed quantification approaches are illustrated in Fig. 12. The first movement uses pathway 5, while the second movement uses pathway 2 (see Fig. 4); both are examples of angry movements. Table 3 provides the associated annotations for both the CMA and the automated approaches. As can be seen from the table, for the pathway 5 movement, the CMA indicated a Strong Weight Effort and a Sudden Time Effort. This was in good agreement with the first automated quantification approach, while the second approach incorrectly labeled

[2]Principal Component Analysis (PCA) is a statistical procedure for finding the directions of highest variations in a multi-dimensional dataset. In the proposed approach, PCA is used to find the plate of movement by finding the two dimensional plane where most of the movement occurs, and therefore the variance is highest.

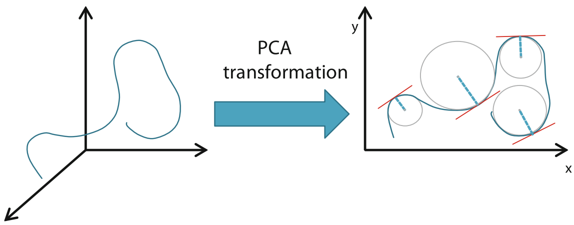


Fig. 11 Quantification for shape directional

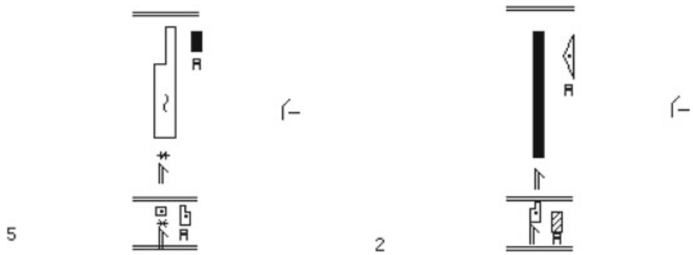


Fig. 12 Motif notation for two example movements as annotated by the CMA. Motif notation provided by Christine Heath

Table 3 CMA and automated annotations for two exemplar movements

	Weight Effort			Time Effort		
	CMA	Automated Q1	Automated Q2	CMA	Automated Q1	Automated Q2
Angry pathway 5	Strong (2)	Strong (2.00)	Light (1.36)	Sudden (2)	Sudden (2.00)	Sudden (1.57)
Angry pathway 2	Strong (2)	Strong (1.09)	Light (0.09)	Sudden (2)	Sudden (1.03)	Sudden (0.73)

In each cell of the table, the label indicates the Effort element (Strong vs. Light for Weight Effort, Sudden vs. Sustained for Time Effort), while the number indicates the annotated magnitude of the element

the Weight as Light. For the pathway 2 movement, the first quantification approach is in agreement with the CMA for Weight Effort (Strong), while the second approach incorrectly labels the movement as weakly Light. Both quantification approaches label the Time Effort as Sustained, in agreement with the CMA annotation.

Considering all the movements in the dataset, for the Weight Effort, high and significant correlation was found between the CMA-annotated and the quantified values, with superior performance being shown by the Q1 approach, as illustrated in

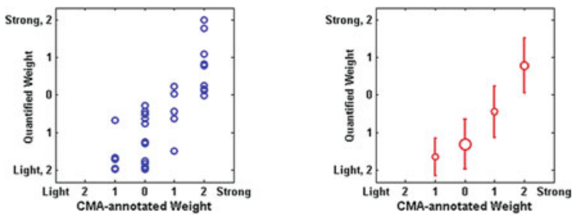


Fig. 13 Correlation between the automated quantification Q1 and the CMA annotation for Weight Effort. The *left panel* plots the quantified Weight and the CMA-annotated Weight for each movement considered. The *right panel* shows the average and standard distribution

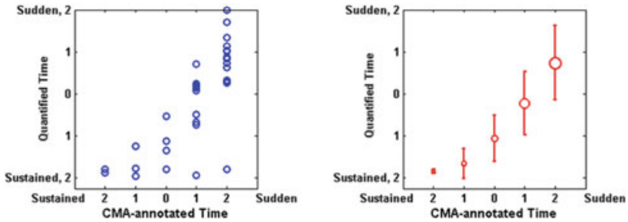


Fig. 14 Correlation between the automated quantification Q1 and the CMA annotation for time effort. The *left panel* plots the quantified Time and the CMA-annotated time for each movement considered. The *right panel* shows the average and standard distribution

Fig. 13. The Pearson correlation coefficient[3] between the Q1 quantification and the analyst ratings for Weight Effort was found to be 81 %.

For Time Effort, the Q1 quantification approach again demonstrated superior results, with a Pearson correlation coefficient of 77 %, illustrated in Fig. 14.

For Shape Directional, a new Boolean measure was proposed, based on a threshold of the Shape Directional quantification, which successfully captures the Shape Directional element when compared to the analyst ratings, as illustrated in Fig. 15. The Phi correlation[4] was found to be 93 %.

For the Space Effort factor, it was difficult to validate the proposed quantifications in the hand-arm movement dataset, due to the imbalance in the sample size, as a large majority of the movements were annotated as Direct by the CMA. Furthermore, Space Effort describes the actor's focus (single-focused vs multi-focused) and other visual cues (eye, head movements) might be needed to better evaluate Space. For instance, an expansive hand and arm movement can be used to greet several arriving parties (multi-focused, Indirect) or a single arriving person (single-focused, Direct), which would be difficult to annotate without additional contextual information.

---

[3]The Pearson Correlation coefficient is a statistical measure of the linear dependence between two continuous variables.

[4]The Phi correlation is a statistical measure of the association between two binary variables.

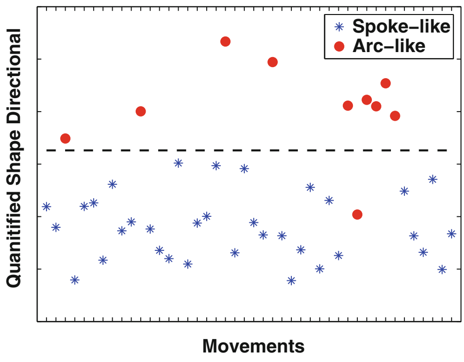


Fig. 15 Automated quantification and the CMA annotation for shape directional. For each movement (arranged along the *horizontal axis*), the quantified shape directional value is plotted on the *vertical axis*. The *dashed line* represents the threshold used to classify each movement as Spoke-Like (below the threshold) or Arc-Like (above the threshold). The CMA annotations for each movement are indicated by the shape of the point, *blue squares* for Spoke-Like and *red circles* for Arc-Like

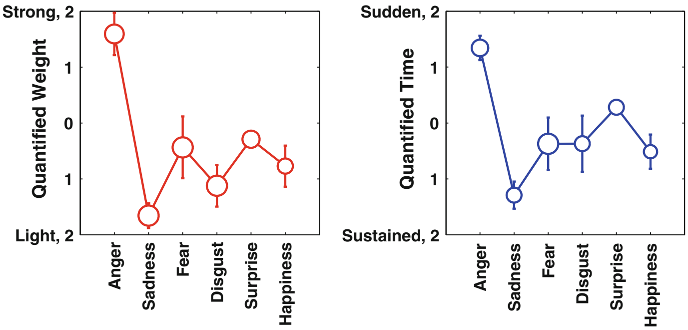


Fig. 16 Average quantified Weight and Time as a function of the emotional category. The *left panel* shows how the quantified Weight varies as a function of the emotion label, while the *right panel* shows the relationship for quantified Time. As can be seen from this plot, emotions are differentiated along these two Effort Factors

For the Flow Effort factor, the correlation between the annotated and the quantified values was found to be 67 %. However, the spatial stopping constraint in the motion paths prescribed to the actor contributes to having movements with multiple Flow qualities, as it turns a movement to Bound Flow toward the end even if it begins as a Free Flow movement.

Finally, the relationship between the quantified Laban factors and the emotional categories of the movement pathways were investigated. Figure 16 illustrates how the quantified Weight and Time factors vary for the movements in different emotion categories. These results indicate that the quantified Laban factors can be used to characterize the expressive content of the hand and arm movements.

7 Expressive Movement Generation

The quantification approach described in Sect. 6 is used within a data-based automated expressive movement generation framework [24]. The quantification outputs, i.e., the LMA factors, are used as a low-dimensional space where similar movements can be more easily found. The goal of the expressive movement generation approach is to imbue a given trajectory with a desired expressive content, in terms of a set of discrete emotional labels. Given a desired motion trajectory, which may or may not contain any expressive content; a target emotion label; and a database of movements with known expressive qualities, the proposed quantification is used to find similar movements (i.e., movements sharing similar Laban Effort characteristics) of the desired emotion class. The similar movements which are identified are then used to train a Hidden Markov Model [25], a stochastic dynamic model which is commonly used for modeling human movement [20]. In this type of model, the movement is modeled as a set of key postures, the spatial variation of each posture, and the dynamics of how one transitions from one posture to the next. The desired movement, consisting of the target movement with the desired affective content overlaid, is then generated using the Viterbi generation approach [25]. In this movement generation approach, the hidden Markov model is used to identify the set of key postures in the model that most closely correspond to the target movement. Then, the posture transition dynamics from the model are used to generate a smooth sequence of postures to produce a movement animation.

The proposed approach is illustrated in Fig. 17. The inputs to the algorithm are: (1) the training movement dataset, consisting of a set of movements for which the associated LMA factor quantification has been computed using the approach described in Sect. 6, and the associated emotion label for each movement; (2) the target emotion; and (3) the desired motion path. Using the LMA factor quantification,

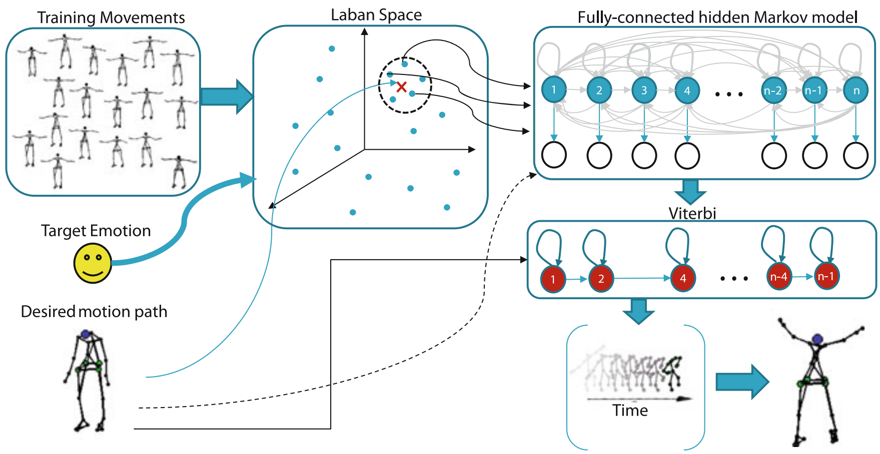


Fig. 17 Proposed expressive movement generation approach

the nearest neighbors (NN) of the target movement with the desired affective label in the database are identified. These movements, together with a number of copies of the target movement, are used to train a Hidden Markov model of the movement. The Viterbi algorithm is then used to generate the most likely key pose sequence in the model for the target movement, and the key pose sequence then used to generate the modified movement. The variable number of copies included in the model is a parameter that can be used to trade off the two goals of the algorithm: increasing the number of copies increases the similarity of the generated movement to the desired motion path, while decreasing the number of copies favors the target emotion over kinematic similarity.

The proposed approach was validated using the UCLIC dataset [26]. The UCLIC dataset is a motion capture dataset consisting of full body movements demonstrated by 13 demonstrators, who each freely expressed movements in 4 emotional categories (sadness, anger, happiness and fear). To test the proposed affective movement generation approach, each movement in the dataset (which had an existing affective label) was used in turn as the desired motion path, and was converted to the other three affective classes. For example, each sad movement was converted to happy, angry and fearful. The generated movements were then evaluated using both an automated recognition approach [27] and by human observers in a user study.

Figures 18 and 19 illustrate two example transformations carried out using the proposed approach. In the first example (Fig. 18), a happy movement from the dataset is converted to a sad movement. In the original movement, the demonstrator's head remains upright while the right arm swings down and across with some Strength and Quickness. In the regenerated movement, the modification results in the head and chest curling forward and down while both arms swing down and closer into the chest with some Strength and Sustainment. In the second example (Fig. 19), a sad movement from the dataset is converted to an angry movement. In the original movement, the demonstrator lowers head, torso and arms down and slightly to the side with some Passive Weight and Sustainment. In the regenerated movement, the modification results in less droop of the head and chest as the hips and legs are engaged and the arms lower to down front with Strength and some Quickness.

Table 4 illustrates the confusion matrix[5] for the automated recognition results. As can be seen from the table, the target emotion is generally correctly recognized, with an overall recognition rate of 72 %, comparable to human perception [27]. Confusions occur most frequently between Fear, Anger and Happiness, categories which share a high arousal level on the dimensional emotion model [28]. Russell [28]

[5]The confusion matrix presents the recognition results in tabular form. Each row indicates the target emotion (the emotion generated by the algorithm), while each column indicates the percentage of time the target emotion was recognized as each category by the recognition algorithm. Perfect recognition would be indicated by 100 % in each diagonal cell. When there are non-zero off-diagonal elements, they indicate what type of error is being made. For example, in Table 4, fearful movement are misrecognized as angry movements 13 % of the time.

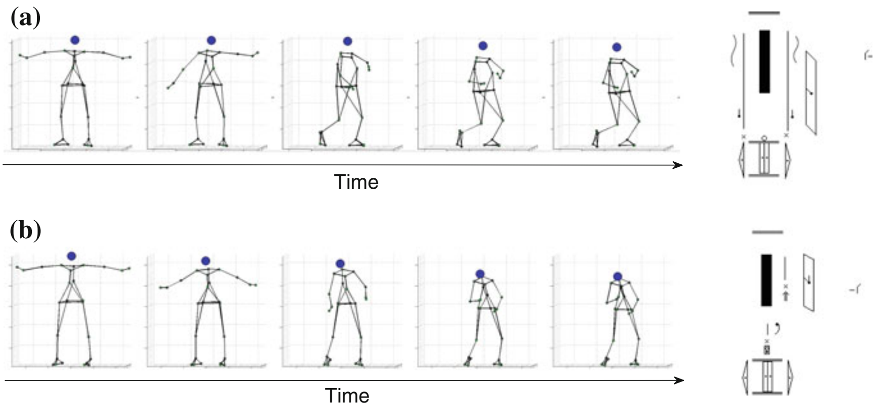


Fig. 18 Example of a movement with an original emotion (happy) re-generated to convey a different emotion (sad). Motif Notation provided by Christine Heath. **a** Original movement (happy), **b** generated movement (sad)

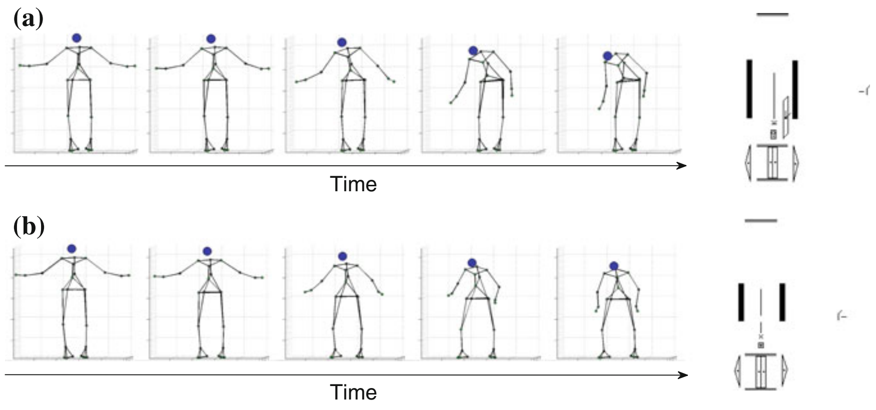


Fig. 19 Example of a movement with an original emotion (sad) re-generated to convey a different emotion (angry). Motif Notation provided by Christine Heath. **a** Original movement (sad), **b** generated movement (angry)

Table 4 Confusion matrix for the automatic recognition of the generated movements

Target Emotions	Recognized emotion			
	Sadness	Happiness	Fear	Anger
Sadness	**83**	1	15	1
Happiness	1	**61**	22	15
Fear	3	8	**77**	13
Anger	1	15	16	**67**

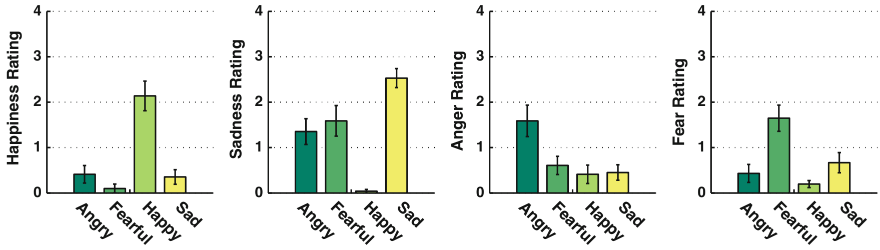


Fig. 20 Participants' ratings when observing animations of the generated movements

postulated that the space of emotions can be represented in a two dimensional space, with the two dimensions consisting of arousal and valence. Arousal indicates the level of activation, alertness or physical activity, while valence indicates whether the emotion is positive or negative. Discrete emotional categories can be mapped to the dimensional model, for example, anger would have high arousal and negative valence, while happiness would have high arousal and positive valence. Fear, Anger and Happiness all share high arousal levels in this model.

Figure 20 illustrates the results of the user study. As can be seen in the figure, observers can generally correctly perceive the target emotion, with the target emotion receiving a significantly higher rating than the other emotions for all the motion types.

Figure 21 illustrates the interaction between the original and the target emotion during human perception of the generated movements. As can be seen in the figure, generated happy movements are perceived as happy regardless of the source movement, while for sad and angry movements, the source fear movement still retains an element of perceived fear. Fear movements could also not be successfully generated from all source movements.

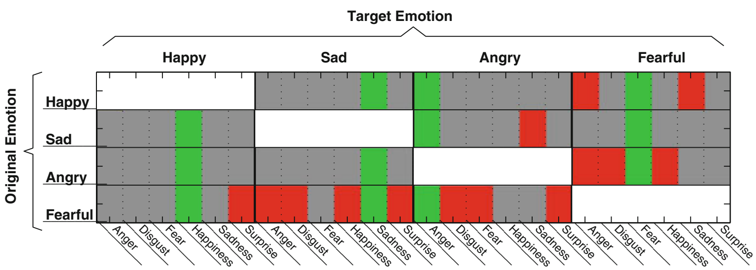


Fig. 21 A heatmap showing significance of pair-wise differences between participants' ratings of target emotions and other emotions (paired t-tests). The *green boxes* highlight the target emotion, the *grey boxes* indicate significant differences to the ratings of the target emotion at $p < 0.05$, and a *red box* indicates that there is no significant difference to ratings of the target emotion at $p < 0.05$

8 Conclusions and Future Work

Laban Movement Analysis offers a comprehensive and concise structure for representing and analyzing expressive movement, which can be of great use for characterizing and generating expressive movement for artificial agents, such as animations, kinetic sculptures and environments, and robots. In this chapter, we proposed an approach for quantifying LMA components from measurable movement features, and using the proposed quantification approach within an expressive movement generation framework. The proposed framework allows movement paths to be imbued with target affective qualities, a first step towards more expressive human-machine interaction.

We are currently working with our collaborators at Philip Beesley Architect to implement the proposed methods in a kinetic sculpture environment, to enable testing with embodied systems and online interaction.

In the future we aim to further explore the other datasets collected, where the hand, fingers and arm are not confined to specific pathways. The knowledge gleaned from further research could be used to help actors and dancers access emotional nuances through guidance and feedback in the process of discovering their own preferences and in expanding their expressive physical vocabulary of movement. Kinetic affective sculptures could be incorporated choreographically into live theatre productions. Also, Sensory Anthropology, a new academic discipline that focuses on how cultures stress different ways of knowing through brain/body maps and the senses [29], might benefit from further investigation of the generation and perception of affective movements.

References

1. R. Gorbet, Personal Interaction, Aug 2010. Venice Biennale
2. B.B. Cohen, *Sensing, Feeling, and Action* (Contact Editions, Northampton, 2007)
3. R. Laban, *The Mastery of Movement*, 3rd edn. (PLAYS Inc, Boston, 1972). (revised and enlarged by Lisa Ullmann)
4. K. Bloom, Interrelationships between between movement analysis and psychoanalysis: a qualitative research project. J. Laban Mov. Stud. **1**(1), 33–43 (2009). Spring
5. B. Kagan, What is spatial tension? Discussion no. 28. http://dnbtheorybb.blogspot.ca/2010/01/what-is-spatial-tension.html. Dance Notation Bureau Theory Bulletin Board (2007)
6. I. Bartenieff, D. Lewis, *Body movement: coping with the environment* (Gordon and Breach Science Publishers, New York, 1980)
7. I. Bartenieff, *Effort-Shape Analysis of Movement: The Unity of Expression and Function* (Albert Einstein College of Medicine, Yeshiva University, 1965)
8. D.M. Chi, A motion control scheme for animating expressive arm movements. PhD thesis, University of Pennsylvania (1999)
9. C. Dell, *A Primer for Movement Description: Using Effort-Shape and Supplementary Concepts* (Dance Notation Bureau Press, New York, 1977). (revised edition)
10. C. Wile, *Moving About: Capturing Movement Highlights Using Motif Notation*. Charlotte Wile Publisher (2010)

11. P. Hackney, *Making Connections: Total Body Integration Through Bartenieff Fundamentals* (Gordon and Breach, Amsterdam, 1988)
12. A. Samadani, S.J. Burton, R. Gorbet, D. Kulić, Laban effort and shape analysis of affective hand and arm movements, in *International Conference on Affective Computing and Intelligent Interaction* (2013), pp. 343–348
13. L. Petit, *The Michael Chekhov Handbook: For the Actor* (Routledge, New York, 2010)
14. G. Stebbins, *Delsarte System of Expession*, 6th edn. (Dance Horizons, New York, 1977)
15. M. North, *Personality Assessment Through Movement* (Macdonald and Evans Ltd., Estover, Plymouth, 1972)
16. K. De Spain, Dance improvisation: creating chaos. Contact Q. **18**(1), 21–27 (1993)
17. P. Ekman, Are there basic emotions? Psychol. Rev. **99**(3), 550–553 (1992)
18. J. Prinz, in *Emotion, Evolution, and Rationality,* Chapter *Which Emotions Are Basic?* ed. by D. Evans, P. Cruse (Oxford University Press, Oxford, 2004)
19. J. Taylor, C. Taylor, *Psychology of Dance* (Human Kinetics, Champaign, 1995)
20. M.E. Karg, A. Samadani, R. Gorbet, K. Kuehnlenz, J. Hoey, D. Kulić, Body movements for affective expression: a survey of automatic recognition and generation. IEEE Trans. Affect. Comput. **4**(4), 341–359 (2013)
21. T. Nakata, T. Mori, T. Sato, Analysis of impression of robot bodily expression. J. Robot. Mechatron. **14**(1), 27–36 (2002)
22. K. Hachimura, K. Takashina, M. Yoshimura, Analysis and evaluation of dancing movement based on LMA, in *IEEE International Workshop on Robot and Human Interactive Communication, ROMAN* (IEE, 2005), pp. 294–299
23. M. Kapadia, I. Chiang, T. Thomas, N.I. Badler, J. Kider, Efficient motion retrieval in large databases, in *Proceedings of the Symposium on Interactive 3D Graphics and Games*, *I3D* (ACM, 2013)
24. A. Samadani, R. Gorbet, D. Kulić, Affective movement generation using Laban effort and shape and hidden markov models. IEEE Trans. Affect. Comput. (2014, Submitted)
25. L.R. Rabiner, A tutorial on hidden markov models and selected applications in speech recognition. Proc. IEEE **77**(2), 257–286 (1989)
26. A. Kleinsmith, P.R. De Silva, N. Bianchi-Berthouze, Cross-cultural differences in recognizing affect from body posture. Interact. Comput. **18**(6), 1371–1389 (2006)
27. A. Samadani, R. Gorbet, D. Kulić, Affective movement recognition based on generative and discriminative stochastic dynamic models. IEEE Trans. Hum. Mach. Syst. **44**(4), 454–467 (2014)
28. J. Russell, A circumplex model of affect. J. Pers. Soc. Psychol. **39**, 1161–1178 (1980)
29. S. Blakeslee, M. Blakeslee, *The Body Has a Mind of Its Own* (Random House, New York, 2007)

Approaches to the Representation of Human Movement: Notation, Animation and Motion Capture

Tom Calvert

Abstract There is shared interest across disciplines in the representation, analysis, composition and visualization of the movement of articulated structures in general and human bodies in particular. Approaches based on different notation systems, keyframe animation and motion capture each have unique advantages but no one offers a comprehensive approach. The unique advantages of the different approaches are discussed together with attempts to combine different approaches. We specifically describe the LabanDancer prototype designed to translate notation into animation, thus providing visualization of the notation. The potential for a truly comprehensive approach is also explored.

1 Introduction

1.1 The Problem

Almost all human activity involves movement—whole body movement for locomotion, manipulation of objects, gestures for communication, aesthetic movement and more. But surprisingly, capturing, representing, recording and analysing movement has proven to be very difficult—largely due to the inherent complexity and multidimensional nature of whole body movement. This chapter explores the elements involved in recording, representing, analysing and visualizing movement of articulated figures in general and human bodies in particular.

T. Calvert (✉)
School of Interactive Arts and Technology, Simon Fraser University,
Surrey, BC, Canada
e-mail: tom@sfu.ca

J.-P. Laumond and N. Abe (eds.), *Dance Notations and Robot Motion*,
Springer Tracts in Advanced Robotics 111,
DOI 10.1007/978-3-319-25739-6_3

1.2 Recording

Recording or capturing movement presents the first problem. The oldest and probably most obvious approach is to capture the essence of the movement in a series of sketches—this has resulted in the development of notation systems that provide a standardized approach. The great advantage of this approach, as will be explored in detail below, is the fact that the sketches or the notation are produced by a trained human who applies analysis at the same time as creating the record. With the advent of technology, video recording has become available. Video provides a relatively inexpensive way of producing a graphic record of movement but it is inherently two dimensional. Creating two videos from different viewpoints can help but does not result in 3D. Stereo 3D video is also available, but does not provide a true 3D representation of the movement that can be viewed from any angle. Motion capture systems provide a variety of ways to generate a numerical description of the moving body that is truly three dimensional and the advantages and disadvantages relative to notation are discussed below.

1.3 Representation and Description

Having recorded or captured movement the question that remains is how to represent it. This has resulted in wide ranging discussions involving applications from detailed hand movements to multiple dancers in a performance. Perhaps the first definitive view since digital technology became available was in the 1979 paper by Badler and Smoliar [1] which concentrated on notation and computer based human figure animation. Now, 35 years later, there is a new and broadly based discussion by Salazar Sutil [2] in his book *Motion and Representation*.

Salazar Sutil considers the representation of human motion through both languages of movement and technological mediation. His argument is that technology transforms the representation of movement and that representation in turn transforms the way we move and what we understand to be movement. To record and capture integrated movement, by means of formal language and technological media, produces a material record and cultural expression of our evolving kinetic minds and identities. This book considers three forms of movement inscription: a written record (notation), a visual record (animation), and a computational record (motion capture). He focuses on what he calls kinetic formalism—formalized movement in such pursuits as dance, sports, live animation, and kinetic art, as well as abstract definitions of movement in mathematics and computer science. He explores the representation of kinetic space and spatio-temporality; the representation of mental plans of movement; movement notation, and such contemporary forms of notation as Choreographic Language Agent [3]; and the impact of digital technology on contemporary representations of movement—in particular motion capture technology. *Motion and Representation* offers a unique cultural theory of movement and of the ever-changing ways of representing movement. The ideas put forward in this book will cause readers to consider movement much more generally than they have before but in some instances it is not clear how this is helpful at a practical level—for example in considering the role of motion capture.

The forms of representation important for this chapter are video, notation, motion capture and figure animation. Video is different from the other modalities in not providing a general representation system. However, video can be useful to overview multiple moving figures and for behavioural studies where researchers wish to document interaction between multiple subjects. Annotation systems have been developed to provide objective records of a 2D video—see ANVIL for example [4].

Notation, as will be discussed below, is a natural way to represent movement. Systems such as Laban notation, for example, can be expanded to any level of detail required. The advantage is that by their nature, notation systems provide a level of analysis but the disadvantages are that notation does not easily allow visualization of the movement and the systems are complex and hard to learn (and thus there are very few trained notators). Motion capture systems, once they are set up, can provide a 3D digital description over time of the movement of key landmarks of the body (typically the joints). This can be used to drive an animated visualization but there is no inherent analysis. With further processing the data can be segmented in time so that the major phases of the movement can be identified and analysed. Segmented motion capture data becomes very similar to human figure animation data. The segments run between keyframes which represent the poses a body takes on when moving. In computer animation interpolation is used to generate in between frames.

1.4 Analysis

Comparing the functionality of notation, animation and motion capture, notation provides the greatest level of analysis and motion capture the least. In contrast, motion capture, which depends on instrumentation, provides the greatest physical accuracy and notation which depends on human judgement provides the least accuracy. In both cases, keyframe based animation is in between. The keyframe represents a pose and the animator will use knowledge of function in refining the movement. This is shown graphically in Table 1.

1.5 Visualization

Whatever approach is used to represent movement, there are many reasons why we wish to visualize the result. The biggest single disadvantage of notation systems is that they do not lead to direct visualization. In contrast both keyframe animation

Table 1 Comparing the characteristics of motion representation modalities

	Level of analysis	Accuracy of representation	Ease of visualization	Digital tools
Notation	High	Low	Hard	LabanWriter Labandancer
Keyframe animation	Medium	Medium	Easy	DanceForms/Life Forms
Motion capture	Low	High	Easy	Many capture systems
Video	Low	N/A	Easy but limited	Any camcorder

and motion capture can be played through an animation system to easily create a 3D animation of one or more animated figures. We discuss below attempts to use notation to drive animation [5]. The main use of a video record is to provide a visualization—but, as noted, the result is 2D and thus is incomplete.

In the remainder of this chapter we will examine the different modalities for representation of movement in more detail, we will discuss attempts to translate between modalities and we will examine differences between these modalities.

2 Notation

2.1 Many Notation Systems

There are many descriptions of the history and evolution of notation systems [6]. Although over 80 notation systems have been proposed or developed only three are in common use: Laban Notation, Benesh Notation and Eshkol-Wachman Notation. We will concentrate on Laban Notation and its extensions. Other chapters of this book provide background and details of different notation systems, specifically Jacqueline Challet-Haas, The problem of recording human movement (Labanotation); Henner Drewes, MovEngine—Developing a Movement Language for 3D Visualization and Composition of Dance (Eskol-Wachman Notation); and Eliane Mirzabekiantz, Benesh Movement Notation for Humanoid Robots?

2.2 Role of the Notator/Choreologist

Whatever notation system is adopted, the notator (or choreologist) plays a crucial role in creating a score. To notate a dance it is necessary to analyse and record the movement in all of its spatial and temporal detail. This is easier with classical ballet, for example, than modern dance; in classical ballet there are many standard poses or movement sequences which do not need to be analysed and notated from scratch each time. At a detail level the notator recognizes the components of a multi-limb movement sequence and notates it appropriately (e.g. By identifying a movement as the flexion of a joint rather than as two adjacent limb segments rotating independently). The notator also recognizes how the movements change over time and in relation to the music.

2.3 Laban Notation and Laban Movement Analysis

Laban notation also known as Labanotation or Kinetography Laban is a notation system developed initially by Rudolf Laban (1879–1958), an important figure in European modern dance (in the remainder of this chapter we will refer to the system as Labanotation). Several people continued the development of the notation and Ann Hutchinson Guest has written one of the classic texts [6]. In the US, the New York-based Dance Notation Bureau supports Labanotation and archives many Labanotation scores (http://www.dancenotation.org).

As illustrated in Fig. 1, symbols are written on a vertical staff where vertical extent indicates time. Vertical columns indicate the body part to which a symbol refers.

Each Labanotation symbol gives four pieces of information. First, the symbol's shape indicates the direction of movement (see Fig. 2a).

Next, the symbol's shading shows the level of a movement, its third dimension; diagonal strokes for high, a dot for middle, and blackened for low (see Fig. 2b).

Third, the symbol's placement on the column on the staff indicates the part of the body that is moving. A Labanotation staff represents the human body; the centreline divides the left side of the body from the right. Symbols to the left of the center line refer to the left-hand side of the body, symbols to the right of the center line to the right-hand side of the body. Some body parts must be identified by a symbol, see examples in Fig. 2c.

Finally, the symbol's length indicates duration of the movement. The staff is read from the bottom up; moving ahead in time (see Fig. 2d). Movements written on the same horizontal line occur simultaneously; movements written one above the other occur sequentially. Measure numbers and dancers' counts appear to the left of the staff.

Labanotation can record movement at a general outline level or can become increasingly specific so that every spatial nuance, dynamic variation, and temporal relationship between individual movements can be clearly stated. Shorthand devices are used by practitioners, but final scores include all necessary detail. A simpler form, named Motif Notation, was developed later by Ann Hutchinson Guest and others as a dance education tool with which both children and adults might explore basic movement actions and concepts. It is also used as a tool for movement observation.

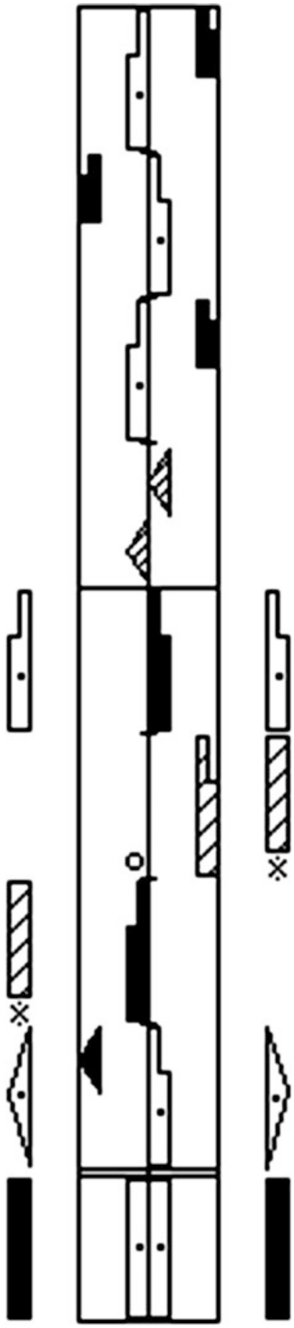


Fig. 1 A simple Labanotation score

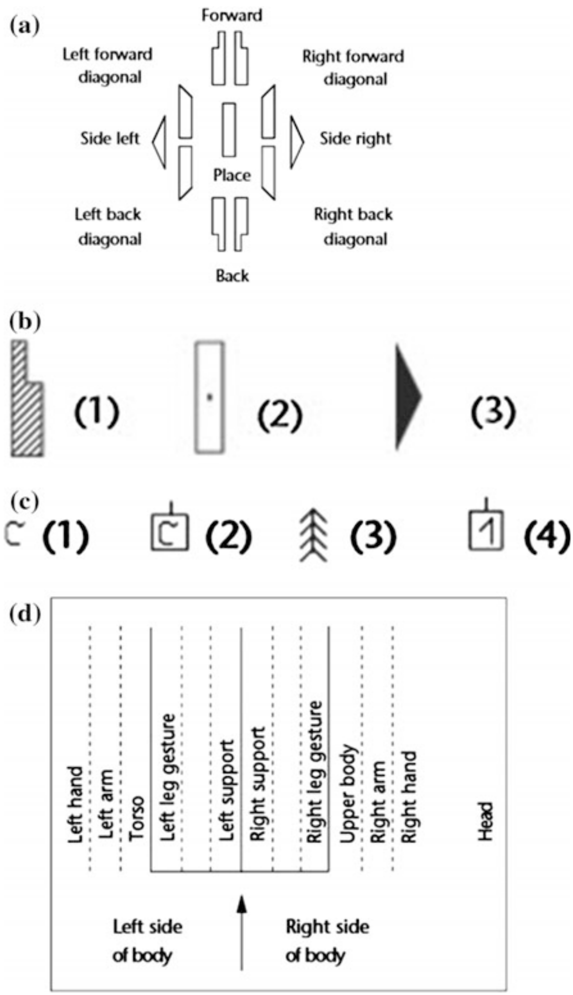


Fig. 2 **a** Direction symbols, **b** Symbol shading: *1* forward high, *2* place middle and *3* right side low. **c** Symbols indicating body parts: *1* head, *2* face, *3* hands and *4* front of left shoulder. **d** The staff

Laban Movement Analysis (LMA) is also based on Laban's work but addresses issues of dynamics and affect not directly addressed by Labanotation. The movement is notated in terms of Effort and Shape using a diagram as illustrated in Fig. 3.

2.4 Gesture and Support

There are three types of Labanotation symbols. Gestures apply to all body parts and map easily to keyframes in an animation system. Support symbols indicate how the body is supported and how the support changes—this does not translate easily to animation. Finally there are a number of symbols used to indicate other aspects

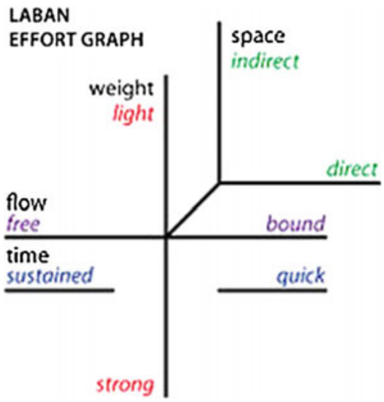


Fig. 3 The Laban effort graph

such as repetition of a sequence, paths, use of floor plans, etc. The result is that in total there are many, many symbols made up of variations and combinations of the main ones.

2.5 Notation Editors

It is natural that as small computers became widely available those working with notation would seek to use the technology to compose and edit scores. LabanWriter was developed at the Ohio State University under the leadership of Lucy Venable starting in 1987. LabanWriter took advantage of the graphics capabilities of the new and relatively inexpensive Macintosh computer to provide a simple and intuitive word processor like system for creating and editing Labanotation scores. The software is available for free download at http://dance.osu.edu/research/dnb/laban-writer.

Other systems for editing Labanotation have been developed. Calaban uses the well-known computer aided design package AutoCAD to edit Labanotation scores (http://www.bham.ac.uk/calaban/) (others have used AutoCAD by itself (http://www.labanatory.com/eng/software.html)). Don Herbison-Evans and his colleagues at University of Technology Sydney have developed the LED editor for selected Labanotation commands [7]. The LED editor interfaces to the LINTER interpreter but this only recognizes a subset of Labanotation. MacBenesh—a similar system for composing and editing Benesh Notation was developed in 1988 by Rhonda Ryman and her colleagues at University of Waterloo [8] and more recently the Benesh Institute has developed an editor for Windows systems (http://www.rad.org.uk/study/Benesh/benesh-notation-editor).

3 Keyframes and Animation

Keyframe based three dimensional computer animation builds on many decades of experience in creating two and three dimensional animations for film and video. The animator creates 2D or 3D keyframes that define the important poses in the movement sequence. In between frames are inserted between keyframes—initially this was done by the animator's assistant and later by a computer interpolation algorithm.

In animation, *a keyframe* is a drawing that defines the start and end points of any smooth transition. The drawings are called frames because they were originally frames on a strip of film. A sequence of keyframes defines which movement the viewer will see, whereas the position of the keyframes on the film, video or animation defines the timing of the movement. Because there are typically only two or three keyframes per second the result does not create the illusion of movement and the remaining frames must be filled with inbetweens.

There are many commercially available computer animation systems that can be used to create and edit keyframe based human figure animation—these include: 3D Studio Max, MotionBuilder, Maya, Unity and more. To simplify the process for choreographers and dancers we have developed a human figure animation system initially called Life Forms and now Danceforms (http://www.charactermotion.com/danceforms). This system is dance friendly and can be mastered much more quickly than the more general purpose animation systems [9]. Perhaps the most famous proponent of using DanceForms for choreography was the late New York choreographer Merce Cunningham, but there are many others. Many schools and colleges have found that this software is also a useful adjunct to teaching dance in the studio. DanceForms and the Ballet Moves II library can be downloaded at no cost from www.charactermotion.com/df-download.html.

The interface used by DanceForms for creating and editing keyframes is shown in Fig. 4. The animator starts with a suitable representation of the body (stick figure, wire mesh, etc.) and adjusts the pose incrementally using forward and inverse kinematics. Joints are selected in turn and spherical potentiometers facilitate setting joint angles.

Each keyframe is placed on the score or timeline as shown in Fig. 5. The software interpolates between each keyframe and inserts "inbetween" frames. The simplest interpolation between keyframes is linear, that is, the inbetween frames change in linear steps between one keyframe and the next. But the interpolation does not have to be linear—a spline curve under the control of the animator can be used to provide any desired profile—an example would be "slow in, fast out" where the velocity of the object in the scene speeds up from the first to the second keyframe. For each keyframe, the figures are placed relative to each other in space using the Stage window (Fig. 6) and the end result, possibly with a metronome or music added is played in the Performance window (Fig. 7).

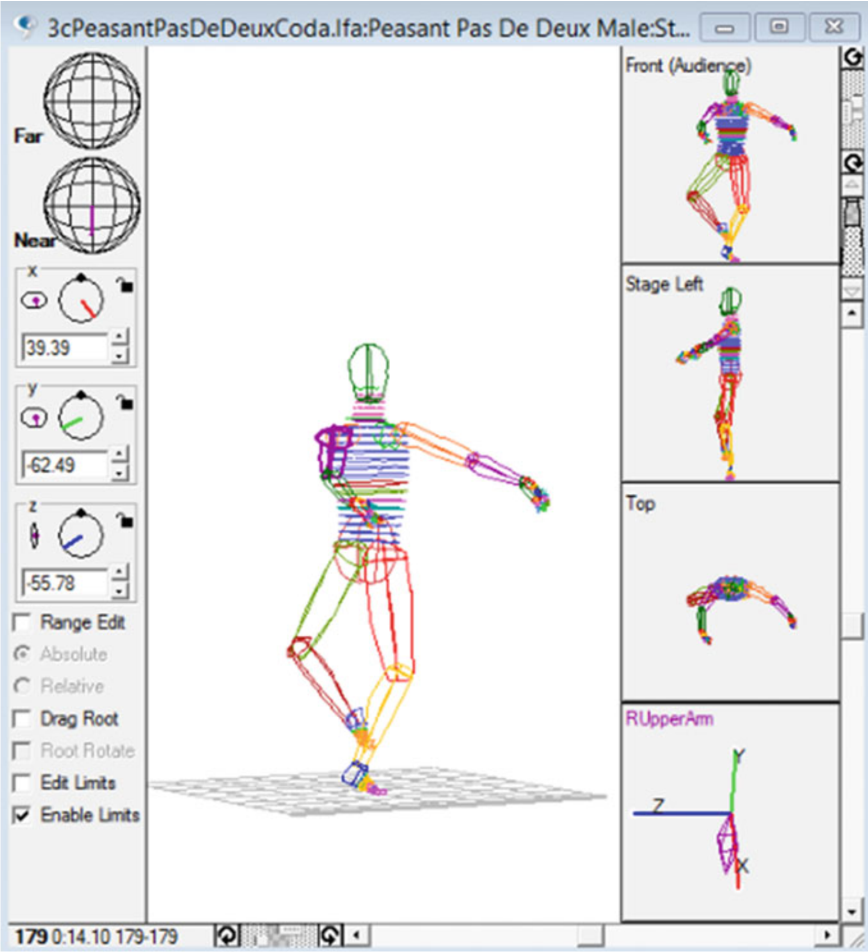


Fig. 4 The DanceForms Studio window is used to compose and edit keyframes. The orientation of each limb segment can be adjusted using spherical potentiometers or by dragging a limb to the desired orientation

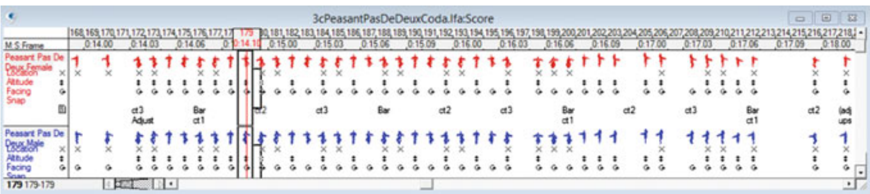


Fig. 5 The score (or timeline) shows the timing of each keyframe. If a keyframe is selected it can be moved in time or edited in the Studio window

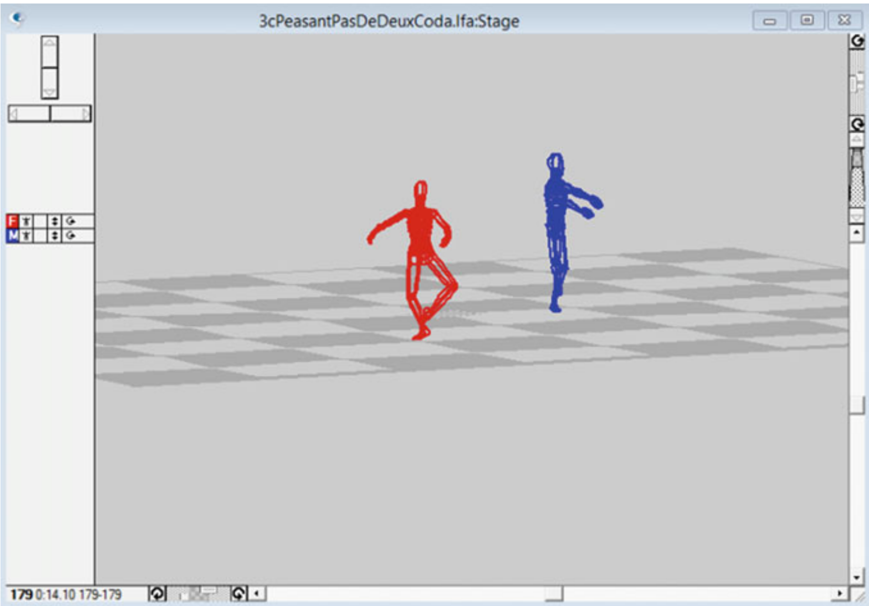


Fig. 6 The stage window allows the figures to be placed relative to each other

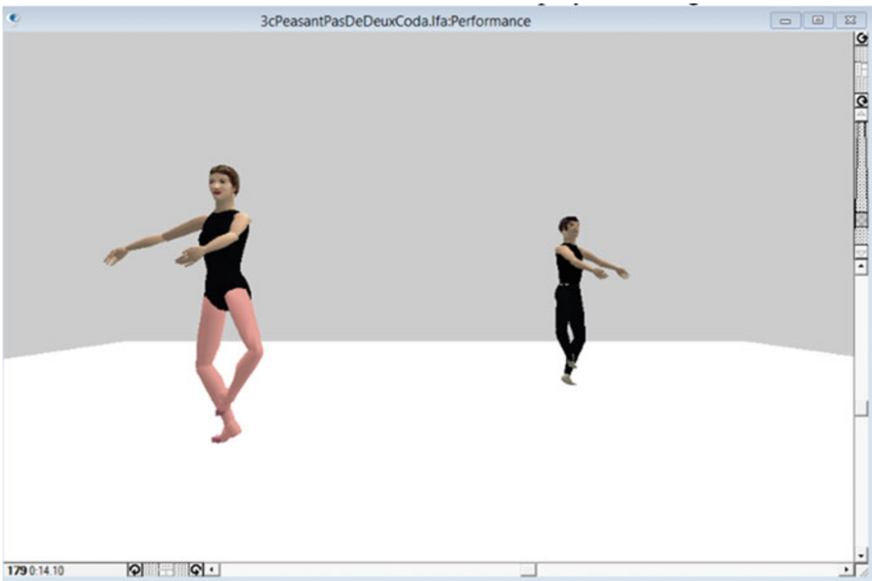


Fig. 7 The performance window where the animation is played through with fleshed out figures

4 Motion Capture

Motion capture (abbreviated mocap) involves the use of technology to capture a digital description of all or part of a moving body. Whatever the technology, however, the goal is to end up with an accurate digital description—this usually means a record of the 3D coordinates in space for each joint of the body for each sampling time.

4.1 Different Technologies

The different technologies available for mocap include:

- Goniometers—these are mechanical devices strapped on the joints to directly measure the joint angles; the instrumentation is cumbersome and inhibits free movement. The devices are more usefully applied to a model of the human body where the limb angles to achieve a particular pose are adjusted by hand.
- Passive optical markers—White markers are attached to the joints of the body and their movement is tracked from different viewpoints by as many as 12 or 18 digital cameras. After initial calibration to identify the markers the cameras track the movement of the markers—hopefully at least 3 or 4 cameras can view all markers at any one time. (Self-occlusion of joint markers by the body cannot always be avoided.) The camera images are analysed by a computer and the location of each marker is recorded at each time step, typically 60 times per second.
- Active optical markers—some of the complexity in mocap with passive markers can be avoided if active markers are used. The active marker is typically a light emitting diode (LED) where each marker is turned on and off in a particular pattern which enables it to be uniquely identified. This simplifies the identification of markers and minimizes the number of cameras that must be used but the end result is the same.
- Markerless optical system—there is a lot of interest in systems where the subjects do not need to wear markers—but this makes joint position measurement more difficult and less accurate. However, a system which can use video images not specifically prepared for motion capture has a lot of attraction. Currently available systems have limitations—for example, the subject may have to wear specially coloured clothing [10].
- Electro-magnetic sensors—another approach which works particularly well for a limited number of limb segments (e.g. arm gestures) involves surrounding the performance space with an electromagnetic field and placing electromagnetic sensors on the joints. The sensor reads the field, decodes it and sends its position in 3D to a computer. The obvious restriction is the need to create a large enough electromagnetic field.

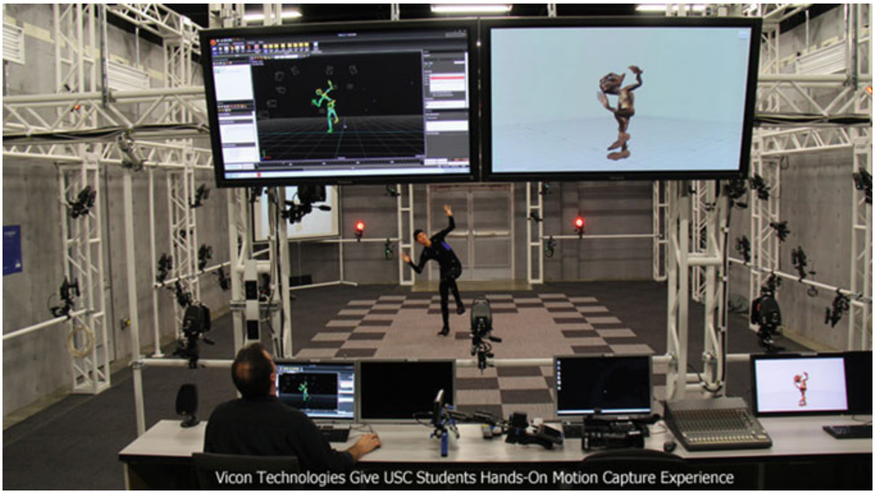


Fig. 8 A typical motion capture studio (Courtesy of University of Southern California http://www.cgarena.com/newsworld/vicon-usc-motion-capture.php)

Large motion capture studios, such as those used by major film and game companies can cost several hundred thousand dollars—a typical example is illustrated in Fig. 8. A more modest installation for a university research lab might cost $20,000 or more. But recently, consumer products like Microsoft Kinect have become available for prices in the range $200–$500. These devices do not replace a full motion capture studio but they do allow small groups or even individuals to capture live movement. There is also potential that in the near future limited motion capture can be achieved using smart phones and tablets which have integral accelerometers.

4.2 Camera Keyframing and iDanceForms

Building on our experience with DanceForms to support the editing, composition and teaching of dance, we have developed a tablet based mobile animation tool, iDanceForms. This new tool takes advantage of the affordances provided by mobile tablets, principally their flexible use in the dance studio. A prototype system implemented on an iPad provides keyframe capture using the tablet camera to create new keyframes. The person posing for the keyframe needs to be dressed in dark (black preferred) clothing in front of a white background. The built-in computer vision algorithm will then capture the pose and search through a database of pre-stored standard poses in order to try to find a corresponding skeleton pose. Once the skeleton pose has been found, it will be added to the list of keyframes. We have designed a database of movement using planar poses that can be easily detected from the front without occlusion. The design was informed by an informal evaluation by dancers and students who showed considerable enthusiasm while using this tool.

4.3 Translation Between Motion Representation Modalities

An obvious problem with different modalities for movement representation is that of translating from one to the other: notation to animation and vice versa, motion capture to notation and motion capture to animation.

4.3.1 Notation to Animation—LabanDancer

Many people have suggested that a computer program be developed to translate notation into animation and a number of prototypes have been developed [11, 12]. The most complete and probably most successful was LabanDancer developed by a consortium of the Dance Notation Bureau, Simon Fraser University, the University of Waterloo and Credo Interactive Inc. [5]. The goal was to use a LabanWriter file as input and create animation similar to that produced by DanceForms.

The first stage of translating LabanWriter files to animation is to use the graphical symbols in the score to create an intermediate representation that organizes symbols according to columns and measures. Here, the graphical symbols are spatially sorted, and assigned to columns and measures based on the symbol position relative to the staff origin. Time dependant symbols, such as direction or turn symbols are first stored in the data structure. Different types of modifier symbols are concurrently collected together and later assigned to the time dependant symbols or columns they modify in a second pass.

Labanotation makes use of modifier signs that are placed below, beside or inside a symbol or column to modify its meaning. Examples include signs for folding and contracting, and signs for specifying body parts. Other symbols are placed next to a sign to indicate that a movement is accented, one part of the body it touching another or a limb is rotated. Other modifiers indicate a value or amount, such as the amount of turn associated with a turn symbol. Path signs are used to modify the path to be followed (e.g. a circular path turning to the right). There are also floor plans to show the positions and paths traveled on stage. Thus the parser that interprets the LabanWriter file first must associate all modifying signs with the symbol they are modifying and deduce from the column whether it is a support or gestural movement. The result of this parsing is a composite score where each element of the score represents a movement. This is not a visible score but rather a time and column ordered list within the software.

A large subset of Labanotation is a direct analog of traditional keyframe animation, that is, explicit time based destination or orientation information is provided by the score which is divided into separate animation channels, in this case represented by the columns. The body model used in the LabanDancer animation system is a deformable mesh model, driven by an invisible skeleton. Each joint in the skeleton is controllable through its own keyframe animation channel. In addition, the arms and legs are controlled by four separate inverse kinematic chains, the goal positions of which are also controlled by keyframe channels. In this way, there is a simple translation for much of the information in the composite score to keyframes in the animation channels. The difficulty occurs in filling in all of the information that is implicit, such as the path a limb must take between keyframes in order to appear natural, or to avoid self-intersection of the model. Context dependent timing changes must also be handled.

The composite score is parsed into three streams. The first interprets gestures—these are non-weight bearing movements of any body part. The second interprets support changes (including locomotion) and the third involves other issues such as repetition of a sequence, paths, use of floor plans, etc. While much of Labanotation is explicit—i.e. it objectively specifies the orientation of limb segments at a particular time, there are numerous instances where the Labanotation is implicit—i.e. it relies on the knowledge of the notator and the reader to deduce the movement in the context of the piece in question. Thus the translator program must include a knowledge base and an inference mechanism to deduce these movements from their context.

Gestures: The Labanotation symbols generally indicate quite unambiguously the start time, end time and final orientation of limb parts involved in a gesture. They do not, however, explicitly specify the path to be followed in carrying out the gesture. The path can be deduced once the starting and end orientations are known. We have implemented an inverse kinematics algorithm and apply constraints to ensure that articulated limbs carry out the movement in an appropriate plane. For some movements it is necessary to add intermediate keyframes as additional constraints to ensure that articulated limbs do not move inappropriately, such as passing through other body parts.

The parser that interprets gestural commands is very simple at a high level—as noted, final limb orientations are usually explicitly specified. However, at a detailed level the parser is extremely complex since it must ensure that every movement is appropriate to the context.

Support: In Labanotation the concept of support and support change is the basis for all locomotion. The notation shows which limb supports the body over a period of time and the direction (if any) of the movement. Thus the notation does not explicitly specify the flexion and extension of the limbs concerned, leaving it to the intelligent performer to recognize which movements are necessary to achieve support change in a specific direction.

The approach that has been adopted is based on van de Panne's footprints algorithm for animation of human locomotion [13]. This algorithm calculates an optimum path for the centre of gravity of a biped based on the placement and timing of the footsteps, and the geometrical configuration. Once the path of the centre of mass is known, keyframes for the feet are generated, based on the footstep position and timing and any restrictions imposed on the flight phases of the foot. The foot keyframes are used to drive the goal positions for the Inverse Kinematics (IK) chains driving the legs. Important in the choice of this algorithm was the fact that it could take changes in level into account, something that is important in Labanotation. Also, the algorithm seamlessly handles the transition from walking, to running, to jumping or hopping, and can handle locomotion along a curved path. With some experimentation we were able to address most of these issues although there will be some movement patterns (often at the limits of what is likely or possible) where the movement may not be as elegant as desired.

One difficulty with the separation of Gestures and Support Changes in Labanotation is that the two sets of commands can provide conflicting animation to limb segments. For example, during a forward walk animated on the basis of support changes, gestures may also be specified for the legs and particularly for the leg that is not providing support. This requires that the gestures be judiciously superimposed on the support changes, but not to the detriment of support.

LabanDancer Prototype: The LabanDancer prototype was implemented for Windows and the Mac OS. The user interface is shown in Fig. 9. There are fairly standard controls that allow the user to adjust the viewpoint by rotating the stage on all three axes and by providing for zoom-in and zoom-out. There is a choice of animated figures—currently male and female ballet and male and female modern. An important feature of the interface is the display of a simultaneous graphic of the

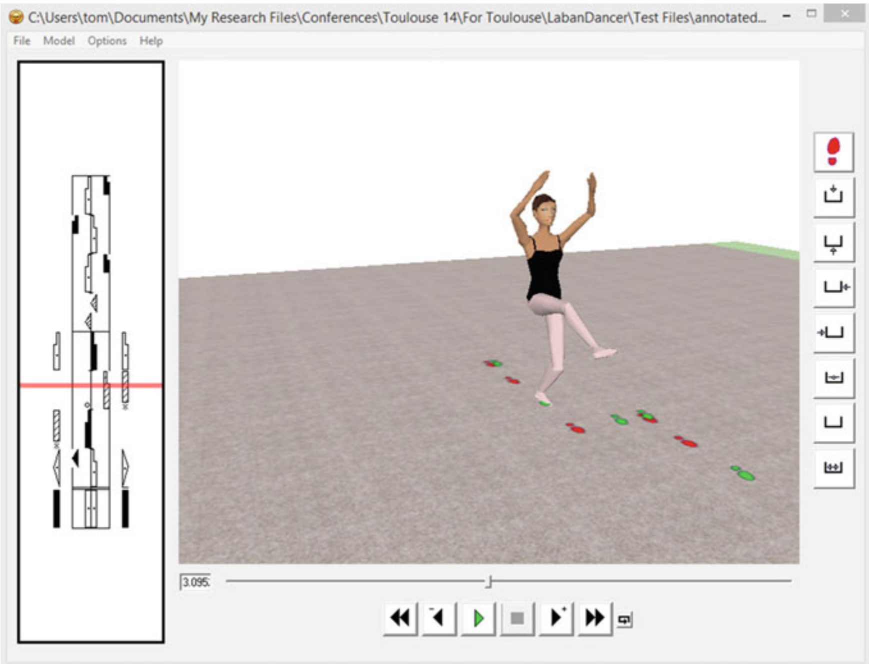


Fig. 9 The LabanDancer interface. The Labanotation score being interpreted is on the *left*

original Labanotation score on the left of the screen. A cursor moves up the score as the animation progresses. While this prototype was quite successful, it is recognized that it is not general enough to handle the very large number of possibilities found in Labanotation. Probably its greatest value has been to clearly identify the structural problems (principally gestures vs. support) found in implementing such a translator. At this time no further development is underway.

4.3.2 Notation to Animation—MovEngine

A rather different approach is being followed in the MovEngine Project under the leadership of Henner Drewes. A notation system based on Eshkol-Wachman Notation and some aspects of Labanotation is used to represent movement and then animated. This is described fully in another chapter of this book.

4.4 Motion Capture to Animation—Thinning

At first sight the translation of motion capture data into keyframe animation should be straightforward. The mocap data is already in the correct format and the only problem is to "thin" the data from (say) 30 frames/s to somewhere between 2 and 10 frames/s. But successful thinning requires that the system identifies frames that are truly "key"—these keyframes should have relatively smooth movement between them. A number of methods have been proposed that look for discontinuities in the movement data for one or more landmarks on the body. To be truly successful, however, human intervention is usually necessary and this can be laborious.

An automatic motion capture segmentation method based on movement qualities derived from Laban Movement Analysis (LMA) has been proposed [14]. LMA provides a good compromise between high-level semantic features, which are difficult to extract for general motions, and low-level kinematic features, which often yield unsophisticated segmentations.

4.5 Motion Capture to Notation

In its full generality, this is a very difficult problem, combining the relatively straightforward thinning problem with the translation of keyframed data into notation. One attempt to do this has been reported by Guo et al. [15]. It is not clear how general the results are. It is relatively easy to achieve acceptable results if the range of movements and symbols is limited.

5 Explicit and Implicit Determination of Movement

Some elements of movement representation, such as keyframes, are explicit and thus under the direct control of the notator or animator. Other elements like interpolation are implicit and only under indirect control. The implications of this are discussed.

5.1 *Poses and Keyframes*

All of the modalities commonly used to represent movement use models of the human body that can be represented by keyframes or their equivalent (we do not consider video since it does not support 3D representation). For mocap and keyframe animation this is quite straightforward: animation is based directly on keyframes and motion capture provides a dense series of frames with embedded keyframes. But for Labanotation it is more complicated. The notation shows which limb supports the body over a period of time and the direction (if any) of the movement. Thus the notation does not explicitly specify the flexion and extension of the limbs concerned, leaving it to the intelligent performer to recognize which movements are necessary to achieve support change in a specific direction. To interpret these support changes with computer algorithms we need to make further assumptions on body dimensions, etc. and to essentially simulate the movements of the lower limbs. The discussion of LabanDancer above outlined one approach to achieving this [5]. As a result of this simulation the movement of the legs can be represented in keyframes. Thus, with this complication, Labanotation, Animation and Motion Capture can all be represented by keyframes that explicitly specify key poses in the movement.

5.2 *Interpolation*

A simple approach to interpolation between keyframes uses a linear equation to calculate the poses to represent the inbetween frames—the typical target frame rate might be 30, 60 or 120/s. There are two potential problems with this. First, if we have a movement path with (say) three keyframes the trajectory at the middle keyframe may not be continuous in the first derivative (velocity) and the second derivative (acceleration)—this can result in a visual "jerk" in the observed movement. Ideally, the inbetween trajectory should be corrected so that the first and second derivatives are continuous [16]. A second possible problem is a situation where we want to change the velocity of the animated object as it goes from one keyframe to the next. For example, this could be a "slow in, fast out" where we want the inbetween frames to have an appropriate profile for an object that accelerates and then adopts a higher speed. This can be achieved by using a spline controlled by the animator to set the value of the inbetween frames. These adjustments to the interpolation may seem complicated and time consuming—in practice they are often ignored and a simple linear interpolation is adopted.

6 Discussion and Conclusions

Many artists and others working with movement share the frustration of trying to find a satisfactory way to represent human movement. On one hand they have notation systems—and that includes Benesh and Eshkol-Wachman notations, not just Labanotation. The great advantage of the notation systems is that in creating a score the notator also analyses the movement and breaks it down into meaningful components. The two major disadvantages are that notation, by its nature, is complex and difficult to learn and that it is hard to visualize the movement. On the other hand, although keyframe animation also involves a level of analysis, it is not as natural as in notation. But keyframe animation has the great advantage that both the individual keyframes and the resulting animation are easy to visualize.

With the premise that there must be a better way, in 2004, 29 interested practitioners and experts gathered at a conference sponsored by the Dance Notation Bureau at the Ohio State University in Columbus, Ohio. The Report from this conference can be found at www.dancenotation.org/news/Ohio_report/Ohioreport.pdf. Following two days of demonstrations and discussions there was consensus that future developments should concentrate on the following points:

1. Build on the capabilities of existing systems rather than develop a completely new system from scratch. This would involve many incremental developments but would probably result in the fastest progress with limited resources.
2. Add a translation capability to different systems so that representations can be easily exchanged (LabanDancer is an example [5]).
3. The new development might be called Interlingua and its capabilities might be best developed in XML since it is easily extensible.

Unfortunately, to our knowledge, lack of resources has inhibited any coordinated follow up on these ambitious suggestions. However, there are steps in this direction—one example is the XML based LabanEditor project led by Hachimura and Nakamura [12] and there is other related research [17]. We hope that this book will help to stimulate and motivate further developments.

Acknowledgement The author acknowledges with gratitude the advice of Marion Bastien (Centre national de la danse) in finalizing this manuscript.

References

1. N.I. Badler and S.W. Smoliar, Digital representations of human movement. Comput. Surv. **11** (1) (1979)
2. N. Salazar Sutil, *Motion and Representation: The Language of Human Movement* (MIT Press, Cambridge, 2015)
3. L. Church, N. Rothwell, M. Downie, S. deLahunta, A.F. Blackwell, Sketching by programming in the choreographic language agent, in *Proceedings of the Psychology of Programming Interest Group Annual Conference*, pp. 163–174 (2012)
4. M. Kipp, Multimedia annotation, querying and analysis in ANVIL, in *Multimedia Information Extraction, Chapter 19*, ed. by M. Maybury (Wiley—IEEE Computer Society Press, London, 2012)
5. L. Wilke, T. Calvert, R. Ryman, I. Fox, From dance notation to human animation: The LabanDancer project. Comput. Animat. Virtual Worlds **16**(3–4), 201–211 (2005)
6. A. Hutchinson Guest, *Labanotation: The System of Analyzing and Recording Movement*, 4th edn (Theatre Arts Books, New York, 2005)
7. S. Hunt, G. Politis, D. Herbison-Evans, LED: an interactive graphical editor for labanotation. Technical report (University of Sydney. Basser Department of Computer Science) 343 (1989)
8. R.S. Ryman, R. Hughes-Ryman, The MacBenesh editor: a "word processor" for Benesh notation. Dance Notation J. **4**(2), 17–26 (1986)
9. T. Calvert, L. Wilke, R. Ryman, I. Fox, Applications of computers to dance. IEEE Comput. Graph. Appl. **25**(2), 6–12 (March–April 2005)
10. S. Corazza et al., A markerless motion capture system to study musculoskeletal biomechanics: Visual hull and simulated annealing approach. Ann. Biomed. Eng. **34**(6), 1019–1029 (2006)
11. Z. Wolofsky, Computer interpretation of selected Labanotation commands. MSc (Kinesiology) thesis, Simon Fraser University (1974) http://summit.sfu.ca/item/3440
12. M. Nakamura, K. Hachimura, An XML representation of labanotation, LabanXML, and its implementation on the notation editor LabanEditor2. Review of the National Center for Digitization (Online Journal) **9**, 47–51 (2006)
13. M. van de Panne, From footprints to animation. Comput. Graph. Forum. **16**(4), 211–223 (1977)
14. Durell Bouchard. Automated motion capture segmentation using Laban Movement Analysis. Dissertations available from ProQuest. http://repository.upenn.edu/dissertations/AAI3328529
15. H. Guo, Z. Miao, F. Zhu, G. Zhang, S. Li, Automatic labanotation generation based on human motion capture data. In Proceedings 6th Chinese Pattern Recognition Conference, CCPR 2014, Changsha (2014)
16. S.N. Steketee, N.I. Badler, Parametric keyframe interpolation incorporating kinetic adjustment and phrasing control. In *SIGGRAPH'85 Proceedings of the 12th Annual Conference on Computer Graphics and Interactive Techniques*, pp. 255–262 (1985)
17. K. El Raheb, Y. Ioannidis, A labanotation based ontology for representing dance movement in gesture and sign language, in *Human-Computer Interaction and Embodied Communication* (Springer, Berlin 2012) pp. 106–117

The Problem of Recording Human Motion

Jacqueline Challet-Haas

Abstract For centuries, various essays have been written that attempted to devise a proper notation for recording dance practices and aimed at preserving them as an artistic heritage. With the increasing interest in human behaviour over the XX century, new trials were proposed to solve the recording of human movements, generally speaking, versus specific dance events. Among these various systems, the so-called «Kinetography Laban/Labanotation», invented by Rudolf Laban (a prominent figure of modern dance in Europe) offers a solution, which will be briefly exposed: built primarily on human behavior in its «ordinary» motion, a natural way, the notion of encompassing space, body and time elements simultaneously is at the heart of its conception. This notation tries to enhance the very basic notion of changing processes as opposed to position writing. Until now it has been widely used mostly in theatrical dance fields as a means of building a dance literature as well as preserving ethnic dances, but also as a tool for research and creation in dance. Its simple but broad basis definitely offers a possible development in many other research domains.

1 Introduction

The question of recording human motion is not new. Artists have tried to capture the essence of movement since the early ages, in drawings, sketches and sculptures. When dance components were eventually codified, the need to keep traces of these set movements emerged.

With the rise of «Modern Dance» in the XX century, which is not based on fixed elements, the need for recording movement became even bigger.

With scholarly requirements applied to research on dance, concerning namely dance folklore, traditional, ethnic and theatrical dances, the necessity of notating the material became paramount.

J. Challet-Haas (✉)
Centre national d'Écriture du mouvement, Crépy-en-valois, France
e-mail: jacqueline.challet-haas@wanadoo.fr

J.-P. Laumond and N. Abe (eds.), *Dance Notations and Robot Motion*,
Springer Tracts in Advanced Robotics 111,
DOI 10.1007/978-3-319-25739-6_4

To find an adequate system enabling the drafting on a sheet of paper the volume of the movement proved not to be so easy.

For centuries, in Europe, and strangely enough in Europe only as far as I know, many attempts were made, which were more or less successful.

A brief survey of some dance notations invented over the last centuries will be briefly laid out from the point of view of the utilized methods.

Among these systems, three of them, namely the Laban system, the Benesh system, and the Eshkol-Wachman system, proposed in the course of the XX century, are still very much alive, used mainly in dance related domains but in other fields of activity too.

2 Various Systems of Dance Notation

For an easier approach, 5 different methods for devising a dance notation system over the last centuries have been selected.

First method: use of letters (verbal language)

Since the XIV–XV century attempts to record the dances were found among the treatises, namely in Spain, France, later in the italien codexes, which consisted mainly in utilizing letters to represent various «steps»: «s» for «simple», «d» for

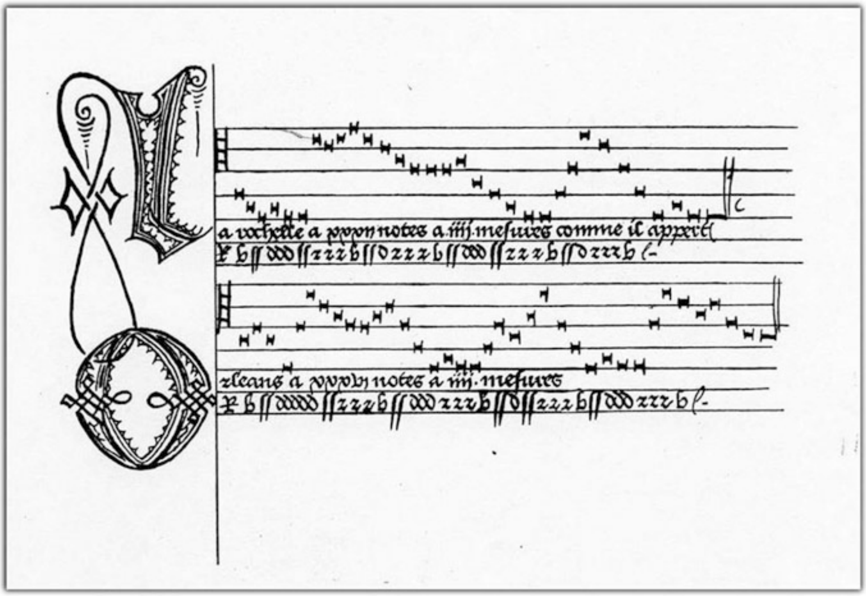
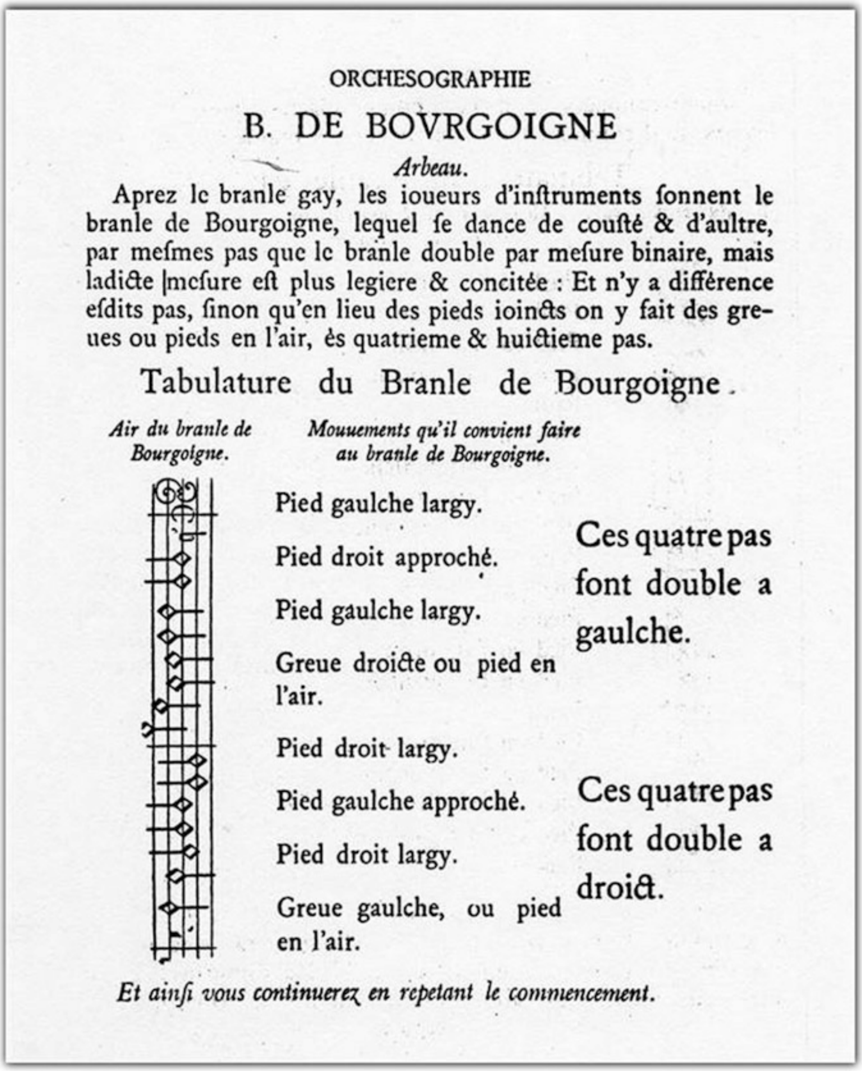


Fig. 1 Manuscript of the Library of Burgundy circa 1450. The succession of abbreviations (*letters*) is written below the musical notes, offering a tentative process to match the steps with the melody

«double», «b» for «branle» etc. providing the fact that these "steps" were already known by the practicians, the conventions being, at these times, strictly established. These treatises served as memory aids (Fig. 1).

At the end of the XVI century, an outstanding treatrise: the "Orchesography" by Thoineau Arbeau, was published in 1588 in Langres (France); it follows the same process but with a definite attempt to match clearly names of steps and music notes and to encode them side by side (Fig. 2).


ORCHESOGRAPHIE

B. DE BOVRGOIGNE

Arbeau.

Aprez le branle gay, les ioueurs d'inſtruments ſonnent le branle de Bourgoigne, lequel ſe dance de couſté & d'aultre, par meſmes pas que le branle double par meſure binaire, mais ladicte meſure eſt plus legiere & concitée : Et n'y a différence eſdits pas, ſinon qu'en lieu des pieds ioincts on y fait des greues ou pieds en l'air, ès quatrieme & huictieme pas.

Tabulature du Branle de Bourgoigne.

Air du branle de Bourgoigne.	*Mouuements qu'il conuient faire au branle de Bourgoigne.*	
	Pied gaulche largy.	Ces quatre pas font double a gaulche.
	Pied droit approché.	
	Pied gaulche largy.	
	Greue droicte ou pied en l'air.	
	Pied droit largy.	Ces quatre pas font double a droict.
	Pied gaulche approché.	
	Pied droit largy.	
	Greue gaulche, ou pied en l'air.	

Et ainſi vous continuerez en repetant le commencement.

Fig. 2 Branle de Bourgogne. The branle de Bourgogne is a circle dance composed of a succession of two «doubles», i.e. 3 steps to the side and a fourth one for closing, circling clockwise and two «doubles» circling anti-clockwise

This use of verbal language can still be found in contemporary dance books, where steps and movements are marked by an abbreviation; many choreographers, dance teachers, and dancers also use this method as memory aids.

Second method: stick figures

"Stick figures" and symbolic drawings were also used as a method for recording dances in a succession of static poses. Among others, the French ballet master Arthur St. Léon in his "Sténochorégraphie" published in Paris (1850), used this method (Fig. 3).

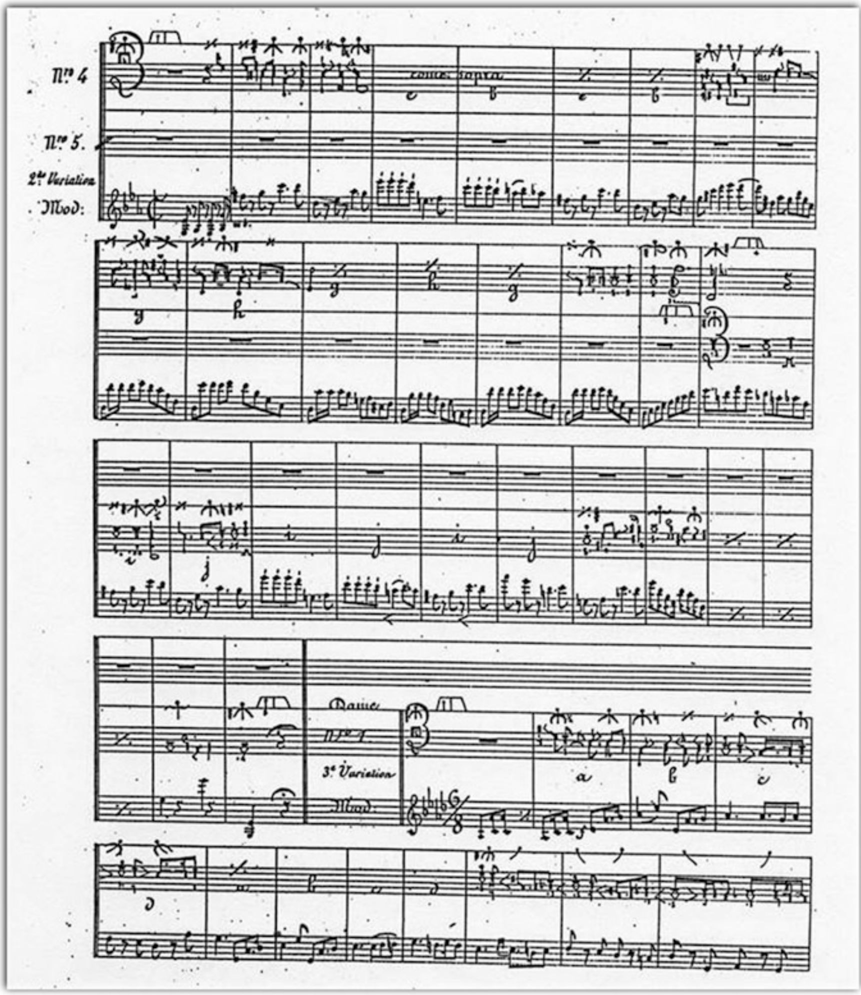


Fig. 3 An extract of the 2nd variation for two dancers from the «Pas de six» of «La Vivandière», chor. A. St Léon. Three music staves are represented: one for each dancer and one for the melody. Inside the *five lines* of the stave, stylised leg gestures are drawn; above, on a subsidiary line, *small figurines* represent upper body movements and arm gestures; on *top*, little «floor plans» appear from time to time

Contemporary choreographers, dancers and dance teachers often draw stick figures as memory aids for their personal use along with verbal notes.

In the same vein, in the middle of the XX century, the system devised by Rudolf Benesh appeared in London in 1956 using this same approach; it follows the display method of accumulating static poses, just as motion-pictures do but devised in a far more abstract way (Fig. 4).

Third method: use of music notes

Another means for recording dances was based on borrowing signs from music notation. At the end of the XIX century, Wladimir Stepanoff from the St. Petersbourg ballet developed a system of dance notation using music notes with some adjustments; it was published in 1898 in Paris under the title: «Alphabet des Mouvements du Corps Humain». Many famous choreographies of the St. Petersbourg ballet were recorded at that time; the great celebrated dancer Vaslav Nijinsky developed this system later and left a notation of his ballet «The Afternoon of a Faun» (Fig. 5).

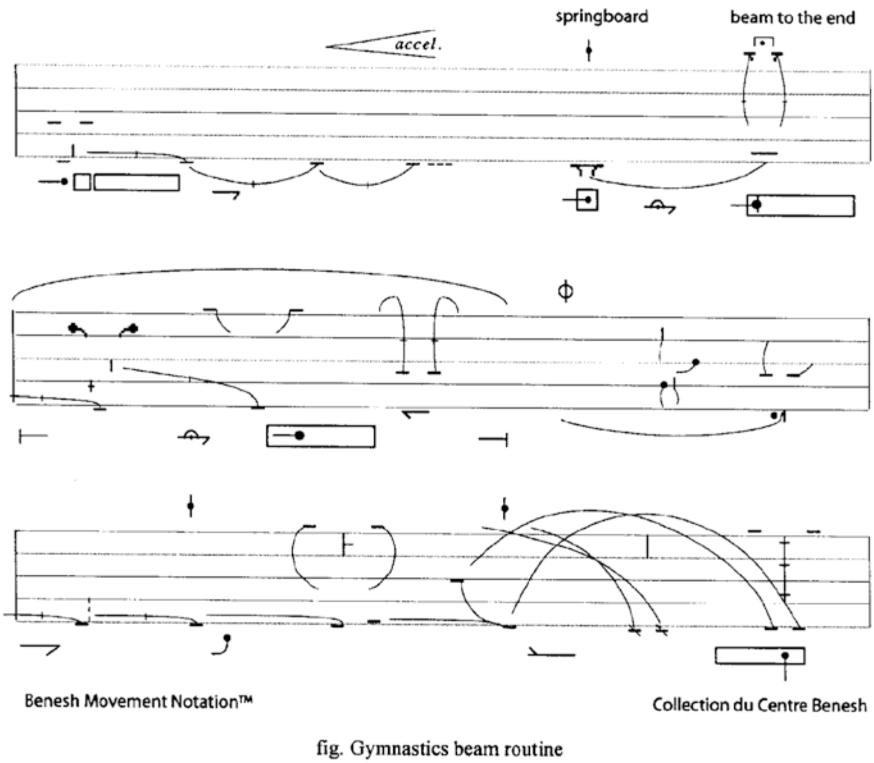


Fig. 4 An extract of a gymnastics beam routine. Points of contact of important moments are written on a stave of five lines dividing the human body in sections; these moments are tied by horizontal bows to show the succession of positions; below the staff is devoted to space indications, above the staff time indications are written (cf. article by Eliane Mirzabekiantz: «Benesh Movement Notation for humanoïd robots»

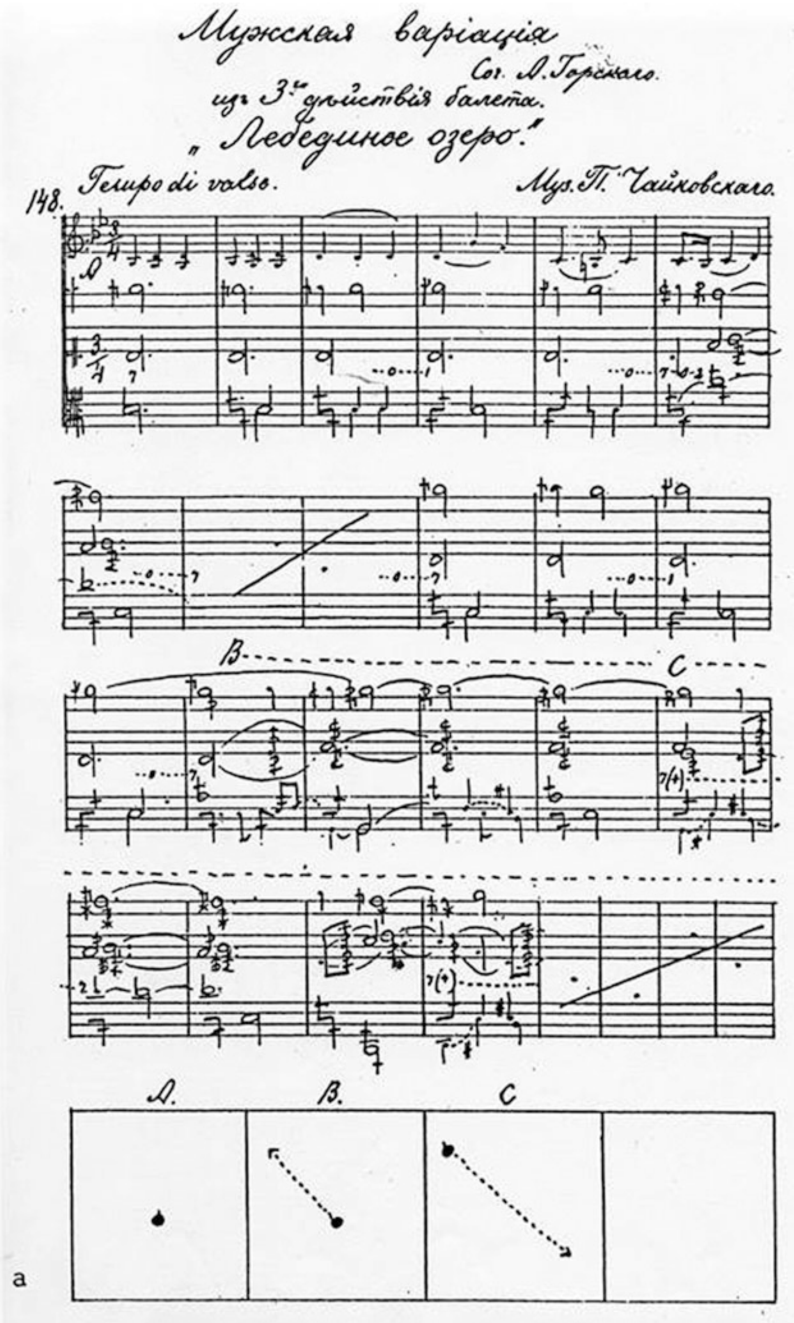



Fig. 5 An extract of «Swan Lake», 2° act, chor. L. Ivanoff. Using music notes on an extended music staff, the time element is incorporated in the so-called “movement signs”; special signs for directions, flexions/extensions, turns are added on the shaft of the music notes, beside or below them

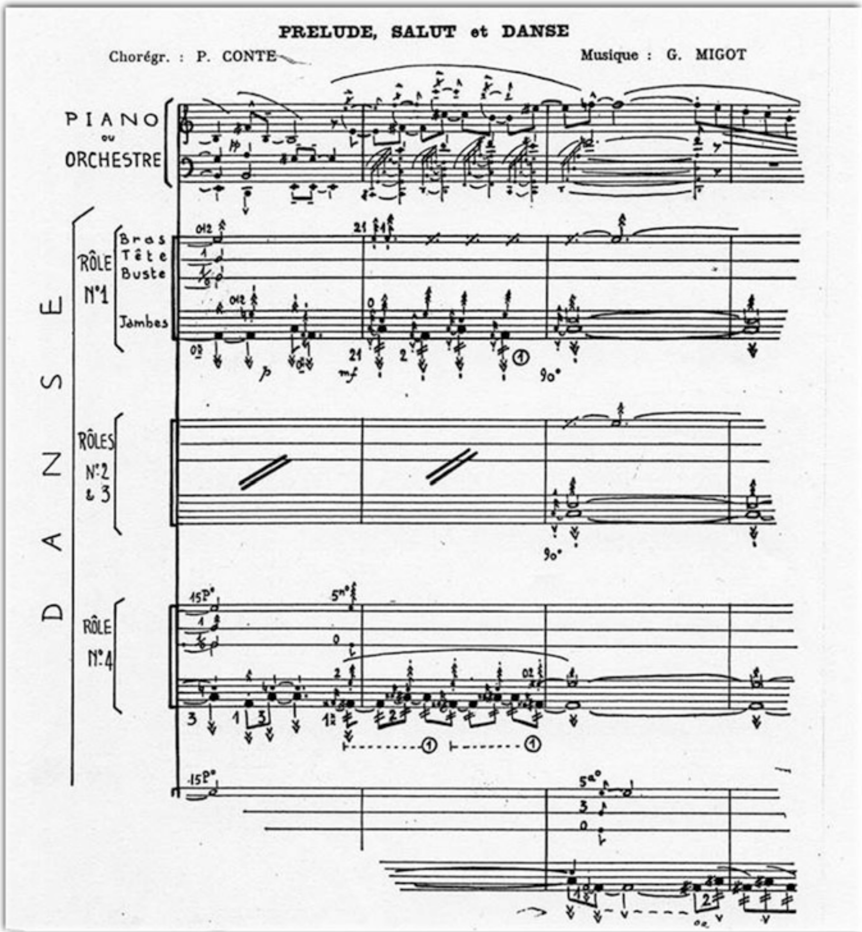



Fig. 6 An extract of «Prélude, Salut et Danse», chor. P. Conté. The music staff is extended: steps and leg gestures are written on and in-between the regular 5 lines, trunk and head movements on *above lines*, arm gestures are on the *top lines*. Space indications are written with *numbers* beside the music notes giving the time element; *small hooks* to indicate flexions/extensions and other specific signs are added at the *bottom* of the music note

This method was later followed by Pierre Conté in his system published in France (1932) and is still in use (Fig. 6).

Fourth method: use of numbers

Another means for recording dance and movement occurrences was proposed by Noa Eshkol and Abraham Wachmann in the middle of the XX century (1958) utilizing then the most abstract symbols, namely numbers.

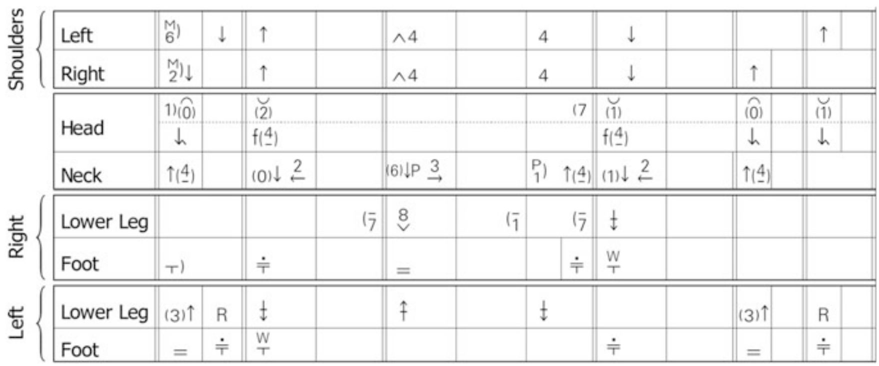


Fig. 7 An extract of «Dove» from the composition «Birds» by Tirza Sapir. The score is read from left to right; *horizontal spaces* represent the different limbs and their parts which are nominated at the beginning of each space. Positions and movements are analysed along a spherical system of reference, the moving limb being the axis. *Numbers* and *small arrows* are used to indicate the coordinates of each position or movement within the circular form drawn by the moving part of the body. *Other signs* are used for indicating contacts, relationships, flexions/extensions etc. *Vertical lines* delineate time units and *thick lines* represent bar lines

(cf. article by Henner Drewes: MovEngine—Developping a Movement language for 3D visualization and composition of Dance) (Fig. 7).

Fifth method: specific movement signs

In the middle of the XVII century under the impulse of the french king Louis XIV, the ballet master Pierre Beauchamps devised a system published in 1700 by Raoul-Auger Feuillet which encoded simultaneously the floor patterns, abstract symbols for set dance units (pas) with basic actions such as plié, sauté, tourné, etc. and the time element. A first, real dance notation was proposed where appropriate components of movement were taken into account. It was used all over France and Europe during a whole century to disseminate the Court dances composed at the French Court. This produced a lot of records, which are a fantastic source of documentation allowing the re-discovery of the "Baroque Dances" during the second half of the XX century not only in Europe, but also in the United States. But this system, still bound to the stylistic features of that period, disappeared with the French Revolution (1789). An entirely new dance culture emerged at that time and this notation system, devoted to one convention only, became outdated (Fig. 8).

Inspired by this Feuillet-Beauchamps system, Rudolf Laban devised at the beginning of the XX century a notation for the «movement», which he called purposefully «Kinetography»; it was published in 1928 in Vienna and quickly disseminated all over Europe and the Americas (namely through the German emigration during the thirties) and lately in Asia.

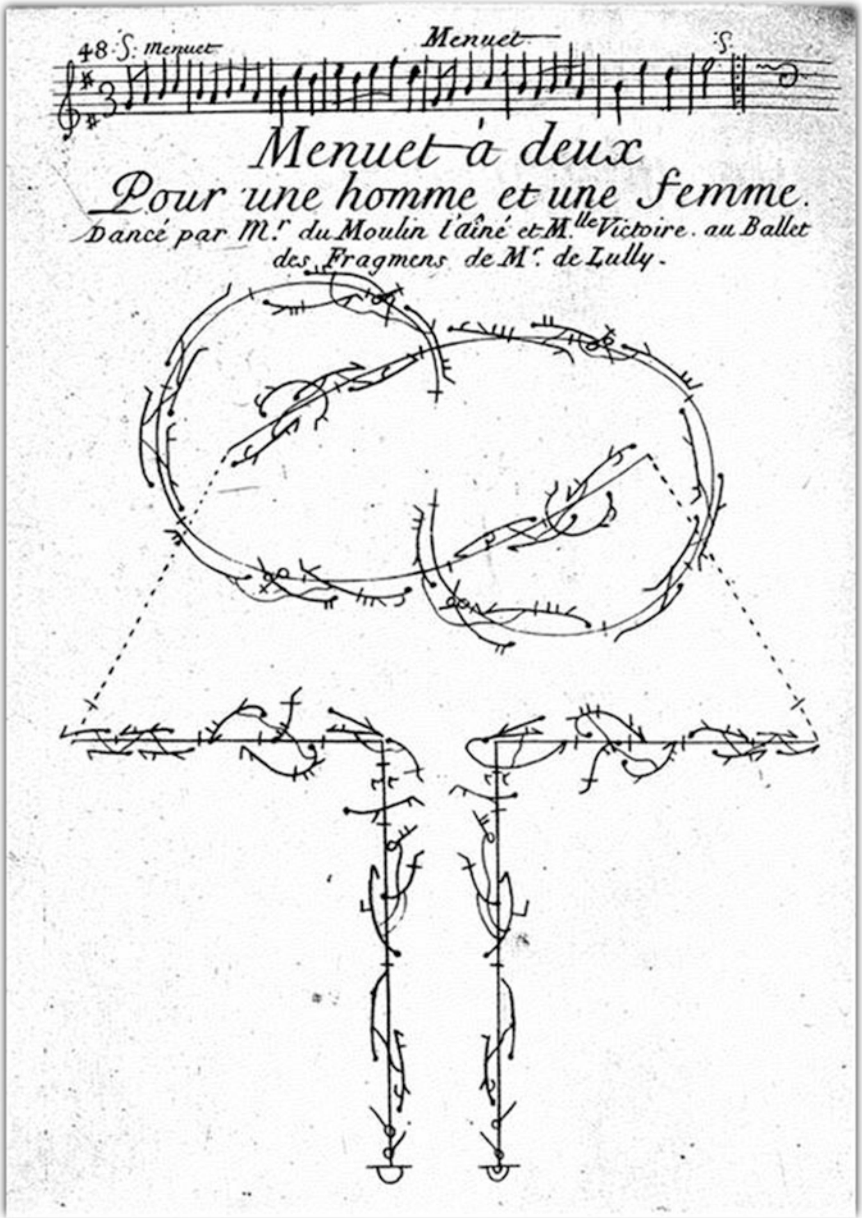


Fig. 8 «Menuet à deux» from «les Fragmens», mus. Lully. The whole page represents the dance area; the *top* of the page, the main audience, the so-called «présence»; two lines of dance, one for each dancer, represent pictorially the design of the dance, the floor pattern; along these lines *tick marks* match the bars of the music, whose melody is written on the *top* of the page; special signs for «steps» are written to the *right* and to the *left* along this line indicating the symmetry of the progression; their various forms indicate the direction of the step, additional small signs attached on the direction signs denote actions such as «plié», «élevé», «sauté»

2.1 Why So Many Attempts?

The question of «why so many attempts» cannot be avoided… The following reasons are suggested:

– the continuing tradition of oral transmission all over the world: the oral, direct transmission functions in various domains, especially where movement skills are involved in dance, play, working activities…. As far as dance activities are concerned, this "functional memory" works quickly and adequately in transmitting established forms of dance in a direct way, often by mimicry. At first glance it seems to be, and is, a quite efficient means of transmission, it does not require any additional effort of externalizing instruction, or any analytical conduct.
– the claim of impossibility to notate dance and movement occurrences in their full display: physically, emotionally; this is primarily based on the difficulty of adequately resolving well established scripts of verbal languages which function in only 2 dimensions?

3 The Movement Notation of Rudolf Laban

Since the beginning of the XX century an interest in the human body and its functions has emerged and developed both in Europe and in the United States.

Rudolf Laban (1879 Bratislava—1958 London), dancer, choreographer, researcher, theoretician claimed for dance an equal status among the "major arts" (music and fine arts). The only way to achieve this goal was to create an adequate notation allowing not only the development of a literature of dance compositions and its dissemination but also offering the possibility to analyse the processes of human movement as he quoted himself in an article published in the magazine «Schrifttanz» (written dance) in July 1928: «the first clear difference which can be perceived in the movement process reveals itself through observing the movement flow (time rhythm) on one hand and the movement shape (space rhythm) on the other».

Laban's search for adequate ways of analysing and notating the components of human movement led him to study the previous attempts of "dance notations" and to propose, with the help of his collaborators, a movement notation published in 1928, called «Kinetography».

It took him around 20 years to conceive it, then another span of 30 years was necessary to achieve a proper development of the system and to be spread and used all over Europe, the Americas and Asia.

His main concern during his lifetime was to observe and analyse how and why human beings are behaving in such different ways in various contexts and their implication in the surrounding world which were developed into 3 domains of investigation:

1. Kinetography called also Labanotation, for analysing and recording human movement processes
2. Choreutik called also Space Harmony for analysing how human beings incorporate the surrounding space
3. Euchinetik called also Effort-Shape for analysing the proper dynamic of each individual.

(cf. article by Angela Loureiro de Souza: «Laban Movement Analysis: scaffolding human movement»)

Kinetography Laban is definitely a movement notation, not a dance notation, like most of the previous attempts; it is not confined to a particular dance convention but based on universal properties of human movement, which are conditioned by:

– the pull of gravity,
– the verticality of the human stance,
– the symmetry of the human body,
– the spatial dimensions.

It is also based on functional issues of human movement:

– the progression of the body in space and time,
– the fundamental distinction between transferring the body weight (changing place) and moving only parts of the body,
– the awareness of his front surface,
– the relativity of the experience depending on context: what precedes affects what is to follow; each mover is unique but similar to any other mover.

To denote these issues:

– abstract symbols are written into a vertical staff, referring to the vertical human stance, which reads from the bottom up of the page as if «walking» on it.
– the staff is divided into columns; the middle line refers to the vertical symmetry of the human body as well as to the line of gravity; movements of the right parts of the body are written at the right side of the middle line; movements of the left parts of the body are written on the left side of the middle line (Fig. 9).
– Each column has a particular destination: the central ones are devoted to the indications of weight transferences (on feet, knee, hands etc.), the next ones to the legs; trunk movements and its parts are written in the third columns; arm gestures are written in the fourth columns; parts of the arms and head movements in extra columns if necessary (Fig. 10). (note: for clarity, only the central columns of the staff are traced).
– the main movement signs are the direction signs, the turn and circular path signs, refering to spatial awareness (Fig. 11).

The direction signs are represented by a rectangle (Fig. 12) which is transformed into 8 main directions: forward, backward, to the right and to the left and the «in-between» ones to the 4 diagonals (Fig. 13).

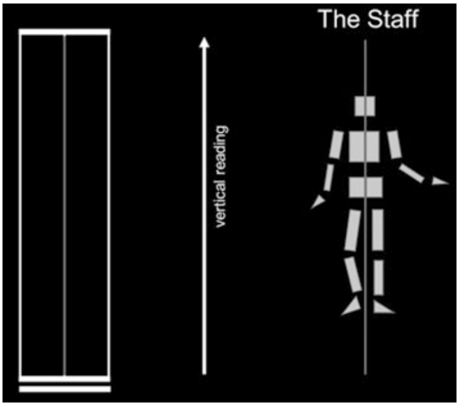


Fig. 9 The staff is divided into columns; the *middle line* refers to the vertical symmetry of the human body as well as to the line of gravity

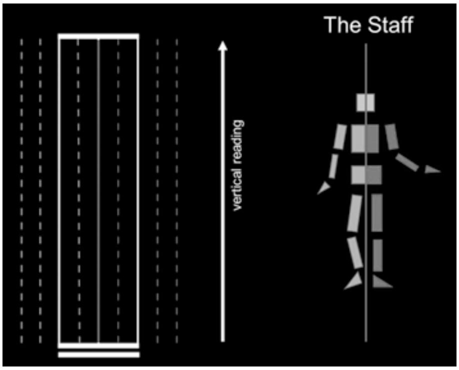


Fig. 10 Each column has a particular destination: the *central ones* are devoted to the indications of weight transferences

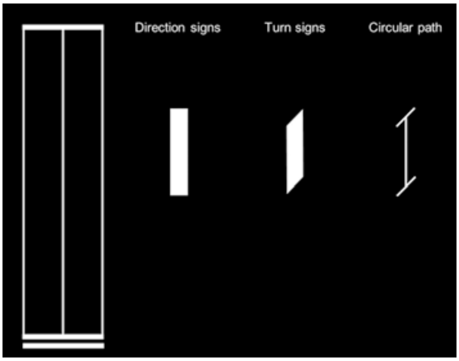


Fig. 11 The main movement signs are the *direction signs*, the *turn* and *circular path signs*, refering to spatial awareness

To denote the volume of the movements, these direction signs are colored: with a dot for medium level, stripped for high level, darkened for low level. Thus 26 main directions are available, which indicate movements going from the body outward; the 27th direction represented by the plain rectangle with a dot inside is the central point, «the place», the only convergent direction (Fig. 14).

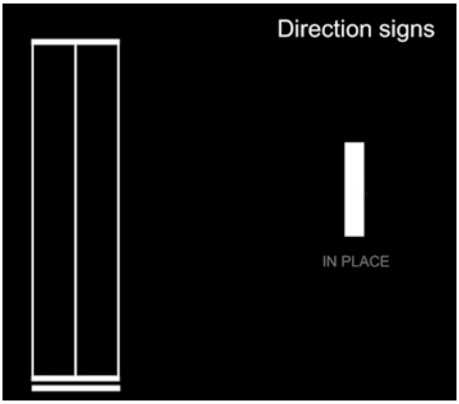


Fig. 12 The direction signs are represented by a *rectangle*

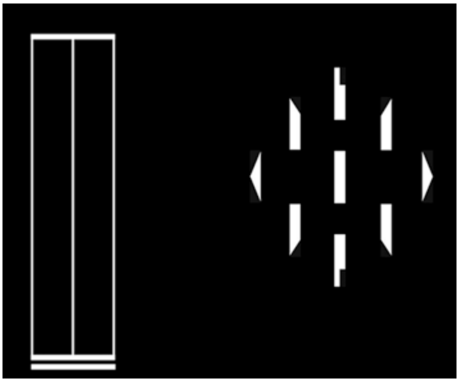


Fig. 13 Forward, backward, to the *right* and to the *left* and the «in-between» ones to the 4 diagonals

The turn signs are derived from the direction signs: the rectangle is cut diagonally at each extremity to denote a clockwise or anticlockwise turn.

The circular path signs are derived from the turn signs: the line is delimited by slanted extremities similar to the turn signs.

These signs have the property of being elongated to denote the duration of the movement, allowing the possibility to encompass in one symbol space and time elements (Fig. 15).

Therefore, one symbol written in the staff represents simultaneously:

- the direction of the movement (by the shape and the coloration of the sign),
- the duration of the movement (by the length of the sign),
- the part of the body moving (by its placement in a column),
- the moment where the movement starts, the moment where it stops (by its location on the staff).

If the simultaneity of space and time indications is certainly one of the main assets of this system, the different treatment of weight transferences as opposed to «gestures» (movements of parts of the body «in the air») is another one. Weight

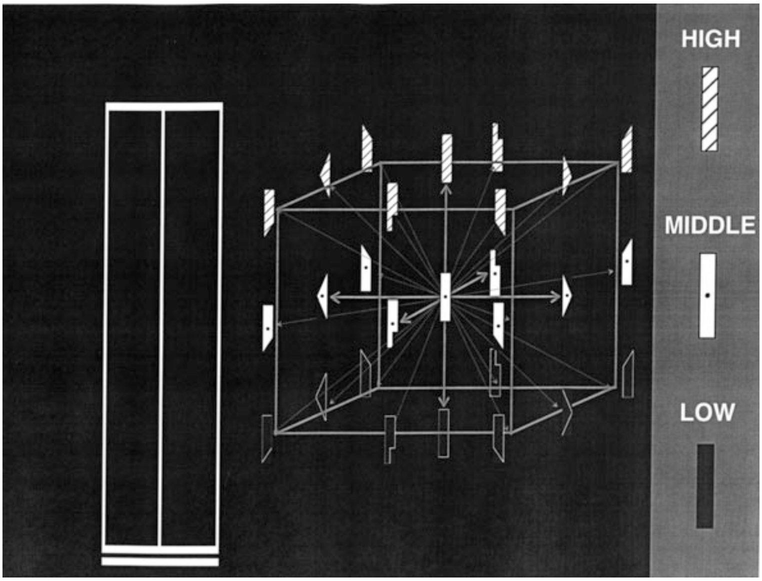


Fig. 14 To denote the volume of the movements, these direction signs are colored: with a *dot* for medium level, *stripped* for high level, *darkened* for low level

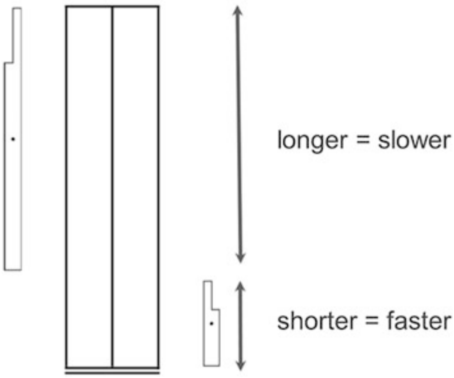


Fig. 15 The *right arm* (the direction sign is located in the 4th column to the right of the central line) performs a movement forward middle (at the level of the shoulder) in one count; then the *left arm* (located in the 4th column at the left of the central line of the staff) performs a similar movement forward middle but in three counts: the sign is longer, the movement being slower. A void in the columns indicates a pause, the cessation or absence of any movement

transferences are treated as proper «motions»: each «step» is relative to the performer; each individual is unique as well as the way he performs his steps; the length of the direction sign encompasses the preparation, execution and completion of the transference but if the size of the step is longer or shorter than the «normal» walking, additional signs will specify it.

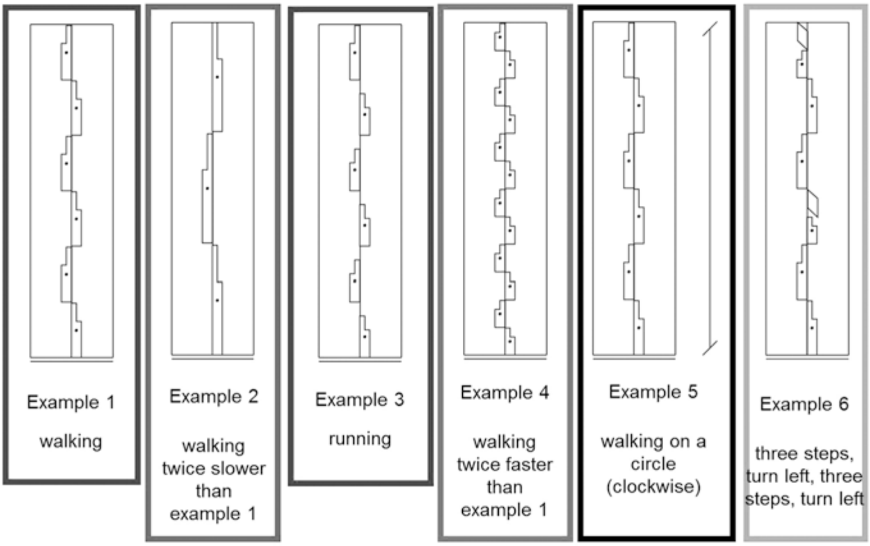


Fig. 16 Simple walking examples

Gestures are treated as «destinations»: each moving part of the body reaches a precise point in the surrounding space from its point of attachment (Fig. 16).

Let us compare two sets of movements. The duration of the six examples is similar:

- 1st example: 6 forward steps, taking one count each, starting on the right foot;
- the 2nd example is composed of only 3 steps, taking 2 counts each: they are twice as slow (but not shorter) than the steps in ex. 1;
- the 3rd example shows 6 running steps: a short gap between the direction signs indicates the cessation of transference of weight, i.e. a small jump;
- in the 4th example, 12 quick steps are written, taking half a count each, twice as fast (not necessarily shorter…) than in the 1st example;
- example 5: a circular path (clockwise) is added outside the staff, the 6 steps of ex. 1 are performed on a curved line; it is a simple way to record a complicated action;
- example 6: the 3rd and 6th steps are turning steps: a turn sign is added above the direction sign.

As a result, the traces of examples 1, 2, 3, 4 are simple straight lines, but in example 5, the line is curved due to the addition of a circular path sign and in example 6, the line is angular due to the addition of turn signs (Fig. 17).

The feet are together in the starting position, the arms are hanging down.

Note: If the absence of symbols in the support columns indicates a jump, because of the cessation of transference of weight, in any gesture column, a void between the signs indicates the cessation of movement, a pause.

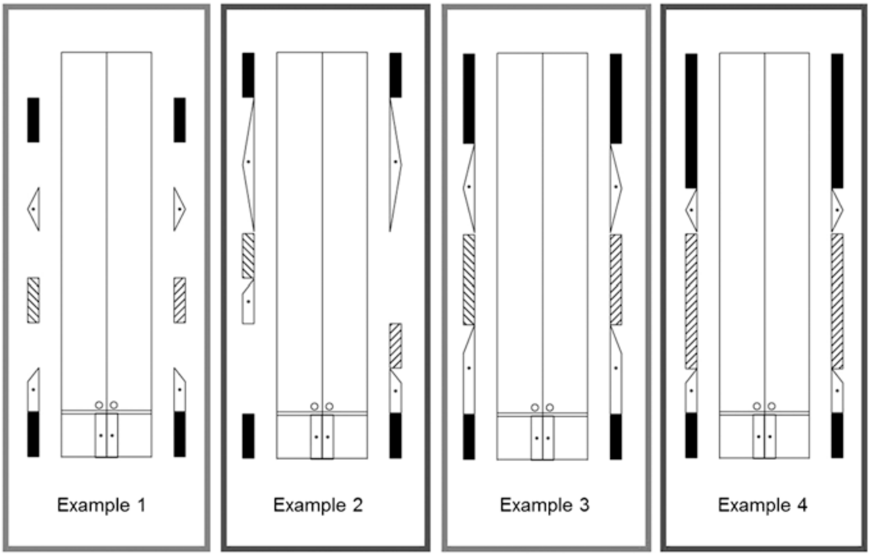


Fig. 17 Simple arm sequences. 4 diagrams of the same arm sequence in 8 regular counts with various movement rhythms

The two «o» in the support columns on the first count indicate a pause of the body to prevent any jump.

– example 1: both arms are lifted together forward and across the front of the body (diagonally) in one count; a pause during one count; both arms go upward during the 3rd count and stay there for one count; on the 5th count, they open to the sides; a pause for one count; on the 7th count, the arms are lowered and return to the starting position; a pause on the 8th count.
– example 2: the right arm goes upward on its own via the left diagonal in 2 counts, then waits until the left arm has reached the upward direction in 2 counts too. Together, they opened slowly to the sides in 3 counts and go quickly down in 1 count.
– example 3: the same sequence as example 1 is performed regularly and smoothly without any stop, each portion taking 2 counts.
– example 4: there is a precipitation in crossing the two arms in 1 count, then a calm elevation in 3 counts, again a precipitation to open the arms to the sides in 1 count and a sustained lowering in 3 counts.

The chosen examples are purposefully simple allowing a clearer overview, but details of level, size of steps, various ways to perform them can be added; flexions/extensions of parts of the body, relationships, dynamic variations beyond the space/time indications etc. may be recorded in details depending on the purpose of the notated text.

The recording of the movement progression encompassing the dynamic element of movement in one symbol, is certainly another main asset of this system.

As it has already been pointed out, an incredibly compact information on the progression of changes in space and time is given at one glance through the incorporation of space and time elements in one symbol and its location on the staff: the symbols written in the central columns indicate the displacement of one's weight in relation to space and to others, as opposed to the same symbols written in the other columns of the staff denoting the destination to be reached by any body part within one's own spatial reach.

Furthermore, not only is the possibility to elicit «motives» offered, allowing comparisons at various levels of investigation, but also the same «sequence» can be written at various levels of complexity depending on the needs. Not to mention the visual aspect of what is happening:

- a «vertical reading» denotes the succession, the progression of changes,
- a «horizontal reading» denotes the simultaneity of the actions of the various body parts, allowing the possibility to get an overview of successive, simultaneous, overlapping actions at one glance (Fig. 18).

A clear exposition of the composition of a whole score is easily perceptible.

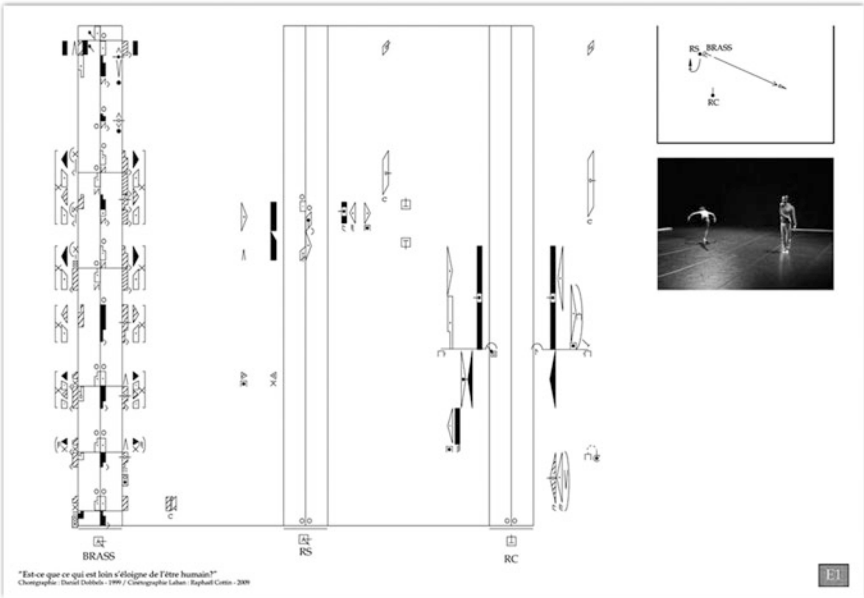


Fig. 18 An extract of a full score: «Est-ce que ce qui est loin s'éloigne de l'être humain» chor. D. Dobbels, 1999, for 3 dancers, notated by Raphaël Cottin, gives an example of a more detailed record

4 Various Domains Using Kinetography Laban/Labanotation

The basic rules of the system are simple but built on a sufficiently broad surface to allow the possibility of developing in various directions, providing the fact that keep the basic principles free from heterogeneous adjustments.

The following domains where Kinetography Laban is already used are:

- Literature, which was one of the first aim of Laban's in inventing a notation system:
  - theatrical dance: a great amount of ballets of various styles has been recorded as well as,
  - traditional, ethnic and sacred dances of various countries,
  - somatic gymnastics, martial arts,
  - mime plays,
  - stage requisites and sets.
- Education:
  Because this notation system allows a description of movement at various levels of complexity: from a very simple, structural form to a most detailed occurrence, it is more and more frequently applied in many educational processes at any age or level of proficiency, in various ways.
  It offers:
  - a better individual understanding of what "one does",
  - a precision of space and interpersonal relationships,
  - a stylistic interpretation in introducing its various components,
  - an access to past and present dance compositions,
  - a means to support improvisation in direct relation to "movement issues" and not on metaphorical instructions only; this led to the development of the so-called «motif notation»; based on structural features of the movement sequence, motif-notation proposes simple diagrams of selected notation symbols in order to enhance the creativity of each performer.
  - in France, namely, an interesting and promising educational project for helping blind people in perceiving themselves in their environment, in developing their kinesthetic appreciation and enhancing their spatial orientation uses kinetographic symbols built in relief (AcaJOUET).
- Research:
  As already mentioned, the Laban script is, since its publication in 1928, in steady development due namely to its spreading over the world but also to the encountering of new movement techniques. To safeguard a proper expansion, the «International Council of Kinetography Laban» (ICKL) was founded in 1961. Since then, biennal conferences are organised offering the possibility to discuss new emergent notational issues, as for instance: the recording of refined

trunk movements, hand movements, floor work etc.… The «motif-notation» is also developing, being incorporated in present dance scores where improvisation becomes an integral part of the choreography.

- Since the Fifties, there is a constant development in methodological approaches in anthropology and ethnology. Kinetography is largely used as an analytical basis for recording and comparing not only dance motives but also specific behavioural habits.
- More recently, research and analyses on choreographic processes are steadily undertaken here and there, thanks to the many available scores.

• Technology:
With the emergence of film, video and computer programs, the question arose: Why still bother with a movement notation? The advantages and disadvantages of film and video will not be envisaged here but the use of computers has had a decisive advantage not only in speeding up the process of transcribing movements but in offering neat scores.
Different programs are available, some semi-official such as *Labanwriter* on MacIntosh computers devised at the Dance Department of Ohio State University (USA) since 1990, regularly updated and available on the web, partly devised by individuals mainly on PC computers.
These programs allow easy copying, publication, dissemination of the material and the introduction of quick corrections, which is in itself a considerable help for notators.
Another program called *Laban Reader* derived from the *Labanwriter* is proposed for educational purpose.
Some other programs dealing specifically with movement analysis are in use and still being developed, such as: *Action Profile* or *Life Forms*.
The flexibility of this system is such that one can easily envisage possibilities to find solutions and adapt its use to other fields dealing with movement occurrences, namely in the domain of humanoïd robotics.

5 Table of Illustrations

Figure 6: Extract from «Prélude, Salut et Danse», chor. P. Conté, Mus. G. Migot, 1931 in «Ecriture by Pierre Conté, Ed. Arts et Mouvement 1931
Figure 7: Extract from «Dove» in «Birds», chor. Tirza Sapir, 2005, Tel-Aviv (courtesy from H. Drewes)
Figure 8: «Menuet à Deux» from «Les fragmens», mus. Lully circa 1700 (pers. Coll.)

Selective Bibliography

Historical Survey

1. Pierre Conté, *Écriture*, Arts et Mouvement, Paris (1931)
2. Noa Eshkol with Michal Shoshnani and Moody Dagan, *Movement Notations, a comparativ study of Labanotation (Kinetography Laban) and Eshkol-Wachman Movement Notation* The Movement Notation Society, Tel Aviv (1979)
3. Raoul-Auger Feuillet, *L'Art de décrire la dance par signes démonstratifs*…Paris, 1700.Ann Hutchinson, *Dance Notation, the process of recording movement on paper,* Dance Books, London (1984)
4. Eliane Mirzabekiantz, *Grammaire de la notation Benesh*, Centre national de la danse, Pantin (2000)
5. Michèle Nadal, *Grammaire de la Notation Conté*, Centre national de la danse, Pantin (2010)
6. Jean Tabourot (Toinot Arbeau), *Orchésographie*, Slatkine Reprints, Genève, (1970)

Kinetography Laban/Labanotation (textbooks)

7. Jacqueline Challet-Haas, *Grammaire de la notation Laban, Cinétographie Laban*, vol. 1–2, Centre national de la danse, Pantin 1999, vol. 3, Centre national de la danse, Pantin (2011)
8. Jacqueline Challet-Haas, *La Symbolisation du Mouvement, issue de la Cinétographie Laban,* Centre national de la danse, Pantin (2014)
9. Ann Hutchinson, *Labanotation,* Theatre Arts Book, 1954, Routledge/theatre Art Books, 4th edn. (2005)
10. Ann Hutchinson Guest, *Your Move, a new approach to the study of movement and dance,* Gordon and Breach Pub. 1983, 3rd edn. (1995)
11. Albrecht Knust, *A Dictionary of Kinetography Laban/Labanotation*, Plymouth, Mac Donald and Evans, 1979, Poznan, 2nd édn. 1997, french edition: *Dictionnaire usuel de cinétographie Laban/Labanotation,* Coeuvres-et-Valsery, Ressouvenances (2011)

Literature: Dance scores (theatre, folklore, ethnic dances)

12. Laban Notation Scores: An International Bibliography
13. Volume 1–4. Warner, Mary Jane with Frederick E. Warner and P. Megan Andrews Ed. ICKL, 1984–1999
14. Notated theatrical dances catalogue, Dance Notation Bureau (DNB), New-York

15. Catalogue des partitions des oeuvres chorégraphiques (notation du mouvement, système Laban), Médiathèque Hector Berlioz, Conservatoire National Supérieur de Musique et de Danse de Paris (CNSMDP)
16. Collection de partitions chorégraphiques, Médiathèque, Centre national de la Danse (CND), Pantin/Paris

Education

17. «Écrire la danse, partager un langage» (an experience in a primary school realised by Bruno Blanche, school teacher and Hélène Leker, notator and dance teacher), DVD, Centre national de la Danse, (CND) (2007)
18. «La partition chorégraphique, outil de transmission, outil d'exploration», (a choregraphic score, tool for transmission, tool for exploration), Noëlle Simonet, Ed: Association Artdir; vol. 1: «Le croquis de parcours» (floor pattern), 2012, vol 2 : «Transferts et tours» (supports and turns) (2015)
19. «AcaJOUET», conception Delphine Demont, a tactile kit for blind people to learning movement issues through movement symbols, Ed. Society Standard non standard, 2008

Research/Anthropology

20. Jean-Michel Beaudet, *Nous danserons jusqu'à l'aube : essai d'ethnologie mouvementée en Amazonie/*: Ed. du comité des travaux historiques et scientifiques (2010)
21. *DANCES STUDIES* Collection since 1976 ed. by Roderyk Lange Centre for Dance Studies—Les Bois St. Peter Jersey C.I., transfered to Instytut Choreologi, Poznan, Poland
22. Farnell, Brenda, and Williams, Drid, *The Laban Script: A Beginning Text on Movement Writing for Non-Dancers*. Canberra: Australian Institute of Aboriginal Studies (1990)

Articles

23. Bertuzzi, Elena, *Différentes modalités d'utilisation de la Cinétographie Laban en Ethnologie de la danse,* ICTM (International Council Traditional Music) Cluj-Napoca (Roumanie), Actes du 24° Symposium ICTM (2006)
24. Bertuzzi, Elena, *Les apports de la cinétographie Laban dans le processus de mémorisation de la danse à travers la partition*—Actes du Colloque «Monaco Dance Forum», Nice (2007)
25. Bertuzzi, Elena*, Oralité et écriture dans la transmission en danse*—Actes du Colloque International «La recherche en danse entre France et Italie: Approches, méthodes et objets» Nice (2014)

MovEngine—Developing a Movement Language for 3D Visualization and Composition of Dance

Henner Drewes

Abstract **MOV**ement-oriented animation **Engine** (MovEngine) is a software library, which was originally developed within a research project conducted at Salzburg University from 2008 until early 2013. One of the objectives in this project was to create a computer application which aids research in re-constructing dance through animated movement sequences, utilizing a movement language based on existing systems of movement notation. Since 2012 the software—in its current developmental stage—is being tested and integrated into movement notation studies at Folkwang University of the Arts in Essen and its development is continued within an MA course in Movement Notation/Movement Analysis at the university. The software allows to create a three-dimensional, animated representation of movement based on a variety of sources and facilitates the exploration of possible variations in the movement material. This new unique and methodologically highly potential technological tool provides the possibility to access referential material on dance and to transfer/translate the *referentiality* into *visuality*, thus revealing the *motoric* and *kinetic* aspects of the material. While the goal of the original project was focused on historic dance research, the employed technical approach may be also applied in a variety of other contexts, e.g. in creating learning tools, in automated animated visualization of movement notation scores or in generating robotic movement. *MovEngine* gains a high degree of flexibility by extending traditional key frame animation techniques with a system of movement orientated instructions, which are based on principles of movement analysis as known from systems of movement notation (*Eshkol Wachman Movement Notation* and *Kinetography Laban*). This paper outlines the key features of MovEngine by describing the role of movement notation principles in the generation of animated movement.

H. Drewes (✉)
Folkwang University of the Arts, Essen, Germany
e-mail: henner.drewes@folkwang-uni.de

J.-P. Laumond and N. Abe (eds.), *Dance Notations and Robot Motion*,
Springer Tracts in Advanced Robotics 111,
DOI 10.1007/978-3-319-25739-6_5

1 Background

The development of MovEngine started in the research project *Visualizing* (*the Derra de Moroda*) Dance Archives at Salzburg University, headed by Claudia Jeschke, who initiated this project together with Henner Drewes in 2008 [1]. The implementation of *MovEngine* is still in progress. During the *Visualizing Dance Archives* project the core functionality of moving and synchronizing free extremities in space according to Eshkol-Wachman Movement Notation (EWMN)[1] principles has been completed. Currently weight transfers, e.g. from one leg to the other, are being implemented allowing the animation of steps and moving in space. The generation of these movements mainly relies on analytical approaches of Kinetography Laban (KIN).[2] Despite its current developmental status with some unfinished features, the advantages of the approach can be already observed in the results of the *Visualizing Dance Archives* project and in the experimental application of MovEngine as a research and learning tool in movement notation studies at *Folkwang University of the Arts* in Essen.

1.1 History

Attempts to digitally process movement notation go back as early as to the late 1960s. Noa Eshkol and her team were invited to the Biological Computer Laboratory at the University of Illinois[3] to create computerized visualizations of the movement paths described by EWMN [4]. The resulting space-chords—as Eshkol called the complex paths created by simultaneously moving limbs (see Fig. 1)—were an early indication of the potential hidden in the analytical approaches.[4] However, this basic research did not yield practical software due to the limited technological capabilities of that time.

[1]Eshkol-Wachman Movement Notation was invented by the dancer and choreographer Noa Eshkol (1924–2007) and the architect Avraham Wachman (1931–2010). It was first published in 1958 [2]. A comprehensive introduction to Eshkol-Wachman Movement Notation will be given in Sect. 2.2 in this article.

[2]Kinetography Laban was introduced by Rudolf von Laban in 1928 [3]. The system is explained in detail in the article *The problem of recording human movement* by Jacqueline Challet-Haas in this publication.

[3]The Biological Computer Laboratory (BCL) was a research institute of the Department of Electrical Engineering at the University of Illinois at Urbana-Champaign. It was founded by Heinz von Foerster, a scientist of Austrian origin, who made important contributions to cybernetics and constructivism. The BCL continued to exist from 1958 until 1874.

[4]New light was shed on Eshkol's artistic work and her early research in the recent exhibition "Sharon Lockhart|Noa Eshkol" by the Los Angeles based artist Sharon Lockhart. See the articles in the catalog of the Vienna exhibition (November 23, 2012–February 24, 2013) [5, 6].

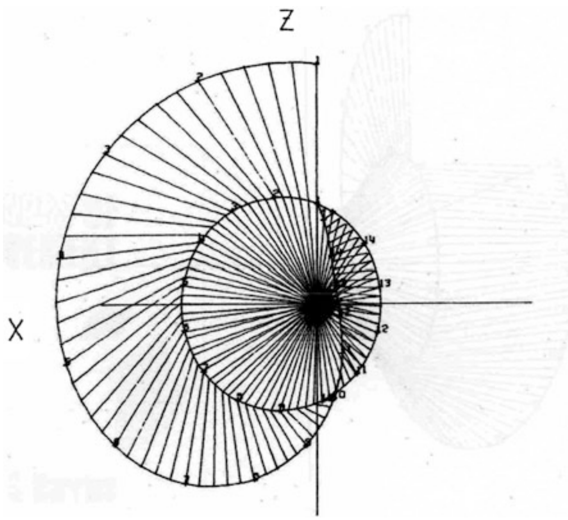


Fig. 1 Space-chords of simultaneously moving abstract limbs [4]

In the late 1970s Smoliar et al. developed a detailed concept to digitally process Labanotation with the ultimate goal to generate animation from digital notation scores [7]. Also later, Labanotation related attempts, which include LED & LINTEL [8] and the LabanDancer [9],[5] project, are based on the objective to ease on the inherent complexity of movement notation and the involved translation processes from the graphical representation to the actual movement, and vice versa. The reading and deciphering of the notation should be automatized to increase the acceptance of notation systems. The user should be freed of the burden to deal with the analysis of movement. Eshkol's work, however, was aimed at the opposite direction: The computerized visualizations should reveal and demonstrate the analytic potential of the notation, and encourage its active usage and application.

The latter approach provides an understanding of the underlying principles of movement notation and creates a foundation for the current animation project. The analytic and abstract methods of movement notation are transferred to visuality to enhance their accessibility, and thus are brought into the focus of users and researchers. Movement knowledge, which denotes an integral and significant, though neglected part of our cultural heritage, may be expressed through a specialized language for movement directly and efficiently. In addition to displaying metaphors for general concepts and ideas on dance, the corporal manifestation of movement can be demonstrated and at the same time its constituting components can be examined. If these components are only partially known, re-composition and re-construction may lead to the original material or to variations of it, allowing to re-experience the kinetic and kinesthetic essence related to the dance. Access is

[5]See also the article *Approaches to the representation of human movement: Notation, Animation and Motion capture* by Tom Calvert in this publication.

gained to the heterogeneous knowledge of dance and its movement content, which can not be codified in ordinary language.

1.2 Aims

The aim of MovEngine is to provide a general, multi-purpose software library to generate 3D character animation, which may aid dancers and researchers in composition, analysis or re-constructing dance through animated movement sequences. It allows to create movement content to be visualized as a visual, three-dimensional representation. The researcher is given a great amount of flexibility, offering a wide range of possibilities and choices to connect visualized body postures to movement phrases and thus helping to construct a realistic representation of a dance.

The software library provides an innovative and specialized 3D animation engine, which has the potential to act as the core in a variety of future visualization tools. It allows to control the resulting animated sequences through centralized parameters. The researcher is able to perform changes dynamically, experimenting with different appearances of the results. This is accomplished by a design which transparently processes dance phrases in a movement-orientated manner. As opposed to traditional, key framed and thus position based animation techniques, this application relies on movement primitives, which build body movement, phrases and ultimately whole dances.

In future developments libraries of dance phrases for different styles will be built, which will function as reference databases for (re-)composing material of certain styles. Previous knowledge on related material will function as a reference, to efficiently reuse standardized patterns in different contexts and variations, to make reasonable choices on how to fill the gaps in the documentation. Visualizing and animating dance material will add an entirely new dimension to composition and research. The increasing transparency of movement material may be considered essential to these developing artistic and academic disciplines, enabling the focusing on underlying movement structures and patterns. Ambiguities of interpretation will become transparent, be resolved and become an independent subject of debate.

1.3 Existing Approaches to Animation

While the graphic animation industry has developed software with amazing capabilities during the last decades influenced by its application in movies and software games, the animation of human movement is accomplished by means that do not easily facilitate restructuring patterns and phrases as necessary when analyzing and composing dance movement. One of the existing approaches is *key frame animation*, which allows for a software-based generation of movement sequences.

Postures of animated figures are distributed over a time line, which are then interpolated by the software to create transitions between the so-called key frames and to produce an animated sequence. In the context of dance, the choreographic software package *DanceForms* [10], which was influenced in its development by the American choreographer Merce Cunningham [11], uses this approach and provides a ready-to-use user interface suitable for choreographers and dancers. Professional animation software exceeds the capabilities of DanceForms by far and allows to control transitions between key frames in greater detail, but also requires expert knowledge to create and edit animations.

All key frame animation approaches, however, are limited in their capabilities to define movement in a comprehensive and refined manner, as movement is basically represented as a series of postures. Transitions may be influenced by different interpolation algorithms, which are not easy to understand for dance animators. Key frame animation does not actively conceptualize movement, as movement is solely defined by transitions between key framed postures. However, conceptualization and categorization of movement is needed to analyze, compose, create or re-assemble dance phrases. All these actions require a proper representation of movement as known e.g. from systems of movement notation.

2 MovEngine and Its Conceptual Framework

2.1 Movement Notation

In contrast to key frame animation, movement notation systems may be regarded as a more suitable and complete conceptual framework to describe movement. They form a basis for accurate description of movement, and serve at the same time as an instrument of thought. MovEngine borrows analytical concepts from two notation systems: Kinetography Laban and Eshkol-Wachman Movement Notation. While substantially different in their graphical appearance and in the application of analytical principles, both notation systems share some basic approaches. Movement is defined as changes in body postures and/or of the location in space. These changes are determined by four analytical components: the part of the body they apply to, the point in time they occur, their duration and the spatial information necessary to reproduce the movement path. Usually, several of such movement instructions must be given simultaneously to describe complex human movement, each for every moving part of the body. Together, they do not only represent the movement of the body as a whole, they reveal the synchronization and coordination of movement.

Generally, both notation systems provide a wide range of alternative descriptive methods for a given movement or movement sequence. Usually these alternatives differ in the way a complex movement is broken down into single instructions and in the granularity of the four analytical components. For example, a bending arm movement may be described in one instruction as one movement of the whole arm, in which the hand is approaching the shoulder. Or, alternatively, the upper arm and forearm movements may be described separately in two instructions. While KIN and EWMN share a considerable amount of these descriptive alternatives, some options are unique to each of the systems. Furthermore, each system defines some default assumptions and a default viewpoint on movement, which influence analytical approaches and notation style. Out of this wide range of analytical options MovEngine first of all selects and uses strategies that allow a most efficient generation of animation. While in the early developmental stages the choices were also determined by programming complexity—the easier ones to implement were chosen first—more alternative and complex options will be added in the future.

2.2 Eshkol-Wachman Movement Notation (EWMN)

EWMN was created and invented by the choreographer Noa Eshkol and the architect Avraham Wachman in Israel and was first published in 1958 [2]. The system is based on an analytical approach, which solely relates to abstract spatial and temporal parameters of movement. The description is based on a limited number of analyzing categories, mainly stating the relations and changes of relation between the various limbs and parts of the body. This information is quantified by numerical values; a fact which facilitates transferring this analytical approach into a digitalized domain as animation. EWMN has been used as a tool for movement and dance creation by Noa Eshkol and her successors mainly in Israel. Also it has become an important component in movement and dance education in this country. Internationally, it gained recognition in a scientific context by researches on neurological syndromes [12] and animal behavior [13]. As it is outside of Israel generally less known than other notation systems as KIN or Benesh Movement Notation [14],[6] a detailed introduction to the system's basics of the system will follow.[7]

2.2.1 Segmentation of Body and Time

The body is seen in EWMN according to its skeletal structure with its limbs moving independently. The movement is captured for each of the moving skeletal sections

[6]See also the article *Benesh Movement Notation for humanoid robots?* by Eliane Mirzabekiantz in this publication.

[7]Many of the primary publications on EWMN by Noa Eshkol include an introduction to the notation system. One of the most complete summaries can be found in *Fifty Lessons by Dr. Moshe Feldenkrais* (1980) [15]. Some more recent additions to the systems are outlined in *Angles and Angels* (1990) [16]. After Eshkol's death in 2007 Sapir and Harries started to summarize certain aspects of the system by comparing Eshkol's primary publications. Their first book *About Time* (2009) [17] mainly deals with aspects of time in EWMN.

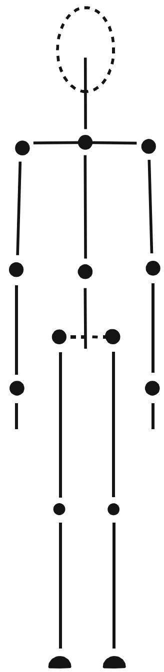


Fig. 2 The body as a system of articulated axes

separately. A part of the body or a limb always lies between two joints or is a free extremity connected to one joint. When analyzing movement, these skeletal sections are imagined as straight lines. The body is seen as a system of articulated axes (Fig. 2).

In a notation score the body is represented through a table-like staff, where the participating moving parts of the body are listed in the left-most column. The succession of the columns represents the flow of time from left to right (see Fig. 3).

The intersections of parts of the body and moments in time form the cells of the score, into which movement symbols are written. Thus, the placement of the symbols denotes the moving part of the body and the point in time a movement begins. The end of a movement is marked by extra vertical lines (bar lines) or by a following movement symbol, which will inevitably cancel the previous one (see Fig. 4).

2.2.2 System of Reference

When circling a part of the body (a movable axis of constant length) in isolation around one fixed end, the other, free end will move along the hull of a sphere. Thus, movement of a single part of the body can be defined as curved paths on the surface of a sphere (see Fig. 5). To describe movement more precisely, a system of

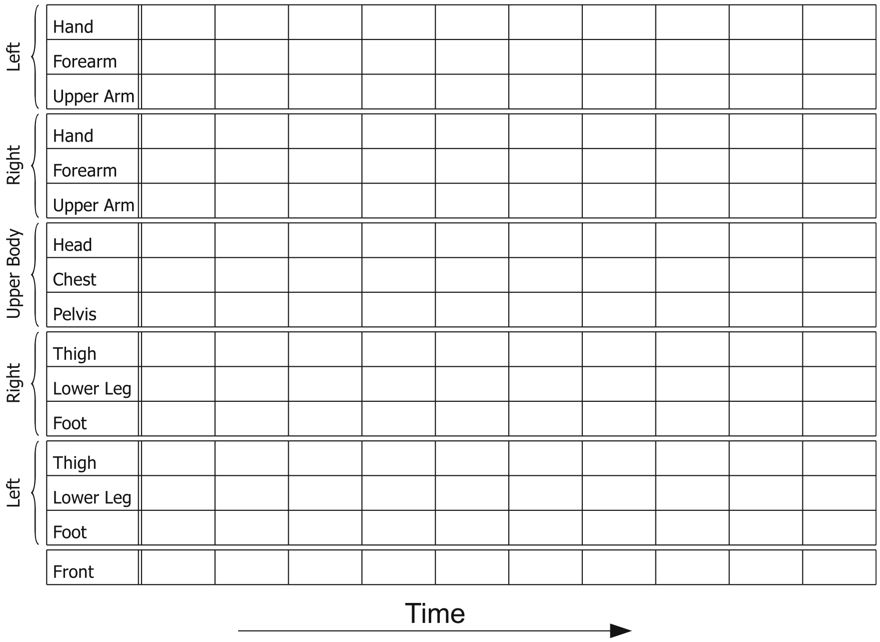


Fig. 3 Scheme of a staff layout

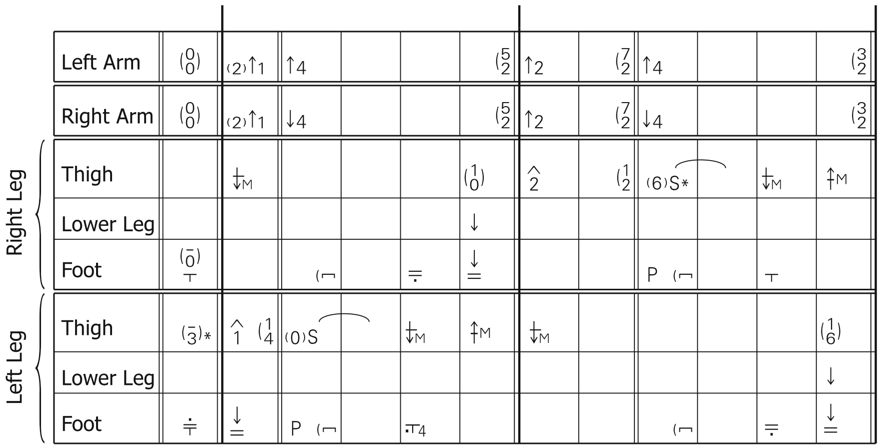


Fig. 4 Example of an EWMN score (excerpt of *Onward and Back*, composed by Tirza Sapir) [18]

coordinates is applied to the sphere. Usually, positions are defined on horizontal or vertical circles (planes) at an interval of 45°, but higher resolutions are possible if necessary in a specific field of application.

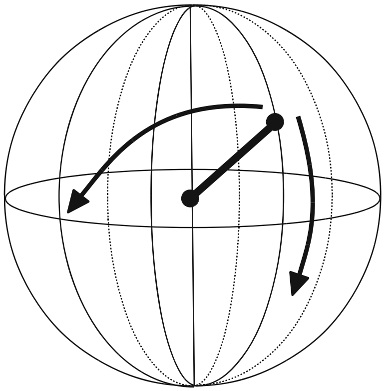


Fig. 5 Spherical movement of a limb

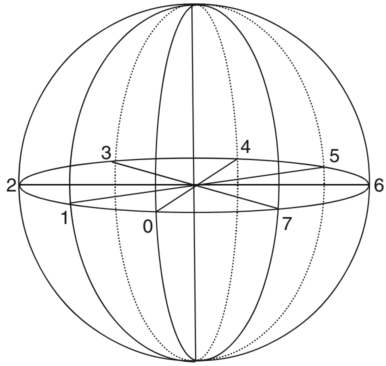


Fig. 6 The horizontal plane

The equatorial plane of the sphere is called *horizontal plane*. When divided according to the scale 1 = 45°, eight positions are obtained on the plane, which are numbered in clock-wise sense from 0 to 7 (see Fig. 6).

The *vertical planes* are perpendicular to the horizontal plane and intersect all with each other at the bottom and the top pole of the sphere. The vertical planes may be identified by the location of their intersection with the horizontal plane. In the scale 1 = 45° four different vertical planes intersect with the horizontal plane at the identified divisions. Each vertical plane is again divided into intervals of 45° and numbered from 0 to 7. The zero position on the vertical planes is set at the bottom pole of the sphere (see Fig. 7).

According to this method 26 positions may be identified on the sphere (system of reference, see Fig. 8). Each position is named using two numbers: The horizontal component denoting the intersection of a vertical plane with the horizontal plane and the vertical component denoting a position on this identified vertical plane.

To describe the state of a part of the body, the center of the sphere needs to be placed at the fixed joint. The free joint or extremity will intersect with the sphere at (or near to) one of the 26 identified positions. Analyzing complex movement of the

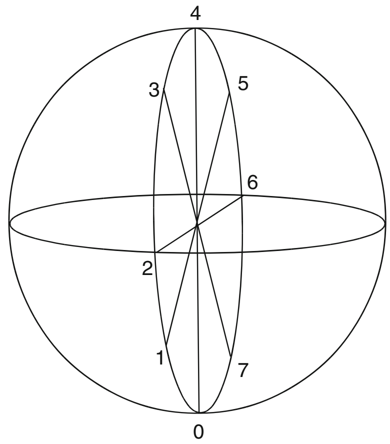


Fig. 7 Example of a vertical plane

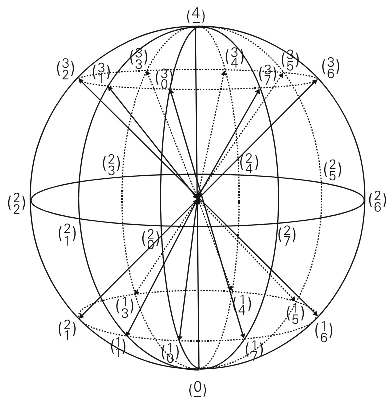


Fig. 8 System of reference with 26 positions

whole human body requires multiple spheres with their systems-of-reference to be located at the joints of all limbs participating in the movement sequence. By default, EWMN uses an *absolute* system of reference with a fixed spatial orientation (global coordinate system). Optionally, a *bodywise* system of reference may be used (see also in Sect. 2.2.6: *bodywise position*), which depends in its orientation on the position of adjacent limbs (local coordinate system).

2.2.3 Axis of Movement and Type of Movement

The movements of a limb are also identified and measured in relation to the system of reference. The curved paths of moving limbs on the surface of their surrounding sphere may be seen as circles or circular arcs. When the limb moves along such a circular path, it circles around an axis called the axis of movement. Since the axis of

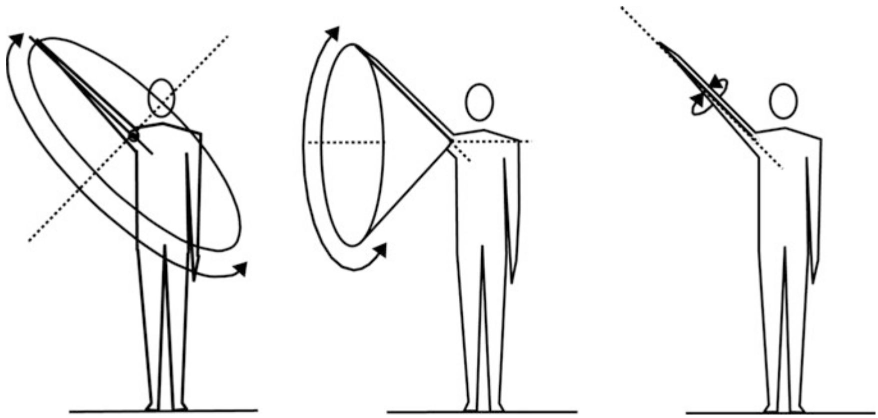


Fig. 9 Three EWMN types of movement: planar, conical and rotational

movement originates at the joint of the moving limb, its position can be defined according to the system of reference.

The *type of movement* is determined by the angular relation between the axis of movement and the position of the axis of the limb at the beginning of the movement. The angle between the two axes establishes the shape of the path of movement and its size. It is possible to differentiate between three different *types of movement* (see Fig. 9):

- *Plane movement*: If the angle between the axis of movement and the limb axis is equal to 90°, the path of movement of the limb's axis creates a plane surface. The extremity of the limb traces the largest circle which can be drawn on the surface of a sphere. Also, connecting any two different limb positions by a movement in the most direct way will always create a plane movement.
- *Conical movement*: If the angle between the axis of movement and the limb axis is smaller than 90° but larger than zero, a conical shape is created by the moving limb's axis. Moving from one limb position to another in an indirect way will create a conical movement. The circles drawn on the surface of the sphere are smaller than the ones traced by a plane movement. Increasing the angle between the limb axis and the axis of movement towards 90° will enlarge the size of the circle drawn on the sphere's surface. Decreasing this angle towards zero degrees will shrink the drawn circle.
- *Rotational movement*: When decreasing the angle between the axis of movement and the limb axis to zero, both axes will be identical. Revolving around the axis of movement will result in a rotation of the limb around its own axis, in which its position in space will remain constant.

Once a type of movement and a specific movement path is established, this information may be further qualified by two additional parameters:

- *Direction*: A limb can travel on the specified circular path in two directions, which are labeled as *positive* and *negative*. For rotational movements this corresponds to a rotation in clockwise and counter-clockwise senses.
- *Amount:* The amount of movement is the angular degree of the taken circular movement path. The information is given in the same scale used in the system of reference, usually 1 = 45°. Example: In a plane or conical movement of eight amounts a full circle is drawn. Applied to a rotation of the whole body a full turn is performed.

Plane movement and conical movement cause a change of the limb in its angular relation to its neighboring limb, while this relation will not change in a rotational movement. Therefore, it is possible to perform a plane or a conical movement and simultaneously combine this with a rotational movement in the same limb.

2.2.4 Law of Light and Heavy Limbs

The description of complex movement involving multiple joints relies on the synchronization of the various simultaneous or successive movements in the parts of the body. To accomplish this, unambiguous guidelines for interdependence and interaction of the individual limbs are needed. The so called *Law of Light and Heavy Limbs* establishes a dynamic hierarchy between the limbs, which is dependent on the supporting body parts (e.g. mostly the feet) and thus it may constantly change throughout a movement sequence.

Any supporting limb in contact with the ground functions as a *base*. From this base a hierarchy is built through all adjacent limbs towards the free moving extremities. A limb closer to the base is called *heavy* in relation to its *light* neighbor being further away from the base. The terms '*light*' and '*heavy*' are purely figurative and are not related to the physical weight of the limbs. The hierarchy from heavy to light will change every time when the base of support is shifted to a different part of the body (see Fig. 10).

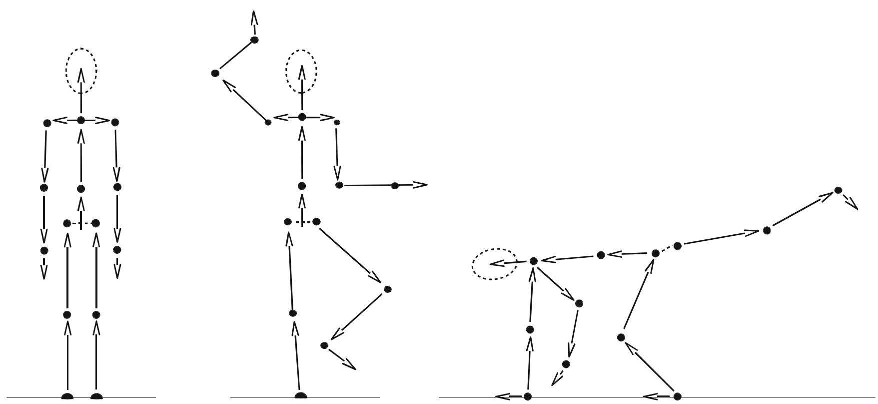


Fig. 10 Examples of changing limb hierarchies according to the law of light and heavy limbs

This changing hierarchy among the limbs implies analytical strategies of how to encode or decode movement to and from the notation. When a heavy limb is actively moving, it will carry its light neighbors to different positions in space. If the light neighbor is simultaneously moving by itself, its movement path will be modified by the movement of the carrying heavy limb. The term '*simultaneous movement*' is used to describe these interdependencies in the articulated linkage of the body. It implies not only the temporal concurrence, but also the creation of the complex paths drawn by the light limbs in space as the product of the circular movements of all participating limbs in the linkage.

2.2.5 Additional Analytical Concepts

The approaches outlined in the sections to represent the foundation of movement analysis in EWMN. However, a full comprehensive movement description frequently requires some additional information. Some complementary concepts relate to the placement and movement of entire limb chains or of the body as a whole. This includes e.g. shifts of weight in the form of steps and jumps or the general front the body is facing to. The state and the changes of the front are expressed by the same means described previously: the system of reference acts as a means of orientation and rotations express the movement around the longitudinal axes of the body. Shifts of weight and steps are usually abbreviated by simply stating the general direction of a step in space in combination with the contacts to, and the releases from the ground. Movement details relating to the single parts of the legs are usually omitted and only stated when necessary. Likewise, it is possible to express general concepts like bending and stretching of arms and legs in a more compact way. Usually, it is more efficient to describe such a movement as one event e.g. of the whole arm or leg rather than as the product of the actions in their single parts (upper arm and forearm/thigh and lower leg).

However, even when employing these general, more compact writing methods, they always reflect the underlying analysis according to the basic EWMN principles for every single limb segment. Even if not efficient in everyday notation practice, each complex movement sequence can be broken down to the circular movement paths in each single limb. This provides consistent analytical means in every possible movement context and enables comparative studies and discovery of movement patterns across a variety of movement styles.

2.2.6 Symbol Usage and Notation Examples

The following table will show some of the most important symbols used in EWMN and introduce their basic meaning. Numerals are replaced by the placeholders x and y. Many of the symbols shown below can also be found in the example score (Fig. 4).

Positions		
$\binom{y}{x}$	Absolute position	Specifies a position in the system of reference. The bottom number ‘x’ refers to a direction on the horizontal plane. The top number ‘y’ refers to the degree on the vertical plane identified by ‘x’. The round parentheses denote the absolute system of reference—the coordinates given are in relation to the surrounding space (global coordinates). Positions are used to specify a start position of a limb (first column of a score) or to state a direct movement of a limb to that position (which usually results in a plane movement)
$\left[{y \atop x}\right]$	Bodywise position	Same as above, but the square brackets denote the bodywise system of reference. Coordinates are given in relation to the neighboring heavy limb (local coordinates)
[ ]	Zero position	The predefined default state for the body and all of its parts is called *zero position.* For most limbs the zero position is considered to be $\left[{0 \atop -}\right]$ (downwards, e.g. arms) or $\left[{4 \atop -}\right]$ (upwards, e.g. parts of the upper body, legs when standing). In zero position the feet are closed to each other and point forward
$\left({y \atop x}\right.$	Mute position	Positions with half parentheses or brackets provide positional information according to the principles explained above, but they do not function as instructions for movement. They may be given at a transitional point or at the destination of a movement mainly as a pure reminder. This is especially useful if the movement defining symbol does not explicitly state a destination position
Movements		
↑ or ↓	Plane movements	General symbol for plane movements. An upwards arrow indicates a positive direction, the downwards arrow indicates a negative direction of the movement. Usually the exact location of the plane in space is given by additional parameters (see below)
(x)↑ or (x)↓	Vertical plane	Movement along the vertical plane identified by (x), either in positive (upwards arrow) or negative (downwards arrow) direction
⟶ or ⟵	Horizontal plane	Movement along the horizontal plane (positive or negative)
∧ or ∨	Conical movement	Conical movement (positive or negative). Additional information may be needed to fully define the location of the movement path
⌒ or ⌣	Rotational movement	Rotational movement (positive or negative)

(continued)

(continued)

x	Amount of movement	A single number without parentheses written above, below or to the right of one of the movement type symbols (plane, cone or rotation) specifies the amount of movement, which is the angular degree of the taken circular movement path
M	Maximum amount	The capital letter M can be given instead of a numerical amount value. The limb should move along its circular path in the largest possible angle. The maximal limit is enforced by the anatomical structure of a joint by other physical constraints
Contacts		
⊤	Ground contact	A limb gets in contact with the ground and serves as a supporting base
⫪	Ground contact without weight	Same as above, but the limb does not carry weight
=	Release contact	A previous contact is canceled
• □	Toes contact or release	A dot above a contact or release symbol applied to feet specifies that only the toes get in contact with the ground or are released from the ground, respectively
□ •	Heel contact or release	A dot below a contact or release symbol applied to the feet specifies that only the heel gets in contact with the ground or is released from the ground, respectively
·⊤4	Contact with rolling from the heel to the whole foot	Used, e.g. in stepping forward, when the heel will touch the ground first
Complex movement instructions involving several limb segments		
(x)S	Step gesture with direction	A stepping gesture, which will involve a bending and stretching of the leg into the given direction (x). To write a full step including the transfer of weight from one foot to the other, additionally the appropriate contact and release symbols need to be written in the foot rows of both legs
↥ or ↧	Stretch or bend	Stretching (upwards arrow) or bending (downwards arrow) a chain of limbs as e.g. the legs. Typically used for the supporting leg

2.3 Kinetography Laban (KIN)

Kinetography Laban was introduced by Rudolf von Laban in 1928 and has continued to develop since then. Most notably this development was lead by Albrecht

Knust [19], and Ann Hutchinson Guest who promoted the system under the name Labanotation [20]. A detailed introduction to the system may be found in the article *The problem of recording human motion* by Jacqueline Challet-Haas in this publication. Therefore, only some major differences to EWMN, which are relevant in the context of creating MovEngine, will be described here.

While the EWMN-based movement analysis defaults to a body image with all single moving parts of the body listed, KIN presents a much more compact view on the body. There are four predefined limb spaces (columns) in the staff for each of the symmetrical sides of the body: support (contact with the ground and transfers of weight), legs, upper body and arms. More limb spaces may be added on demand, either for additional parts of the body (e.g. head, hands and fingers) or for a more precise separation into single skeletal segments. These extra limb spaces are only added when needed. Therefore the appearance of different staff layouts may differ substantially.

This compact view of the body leads to a different analytical approach to movement. When analyzing an arm movement, in KIN the first choice would be to see the whole moving arm, changing its orientation and extension. Only if needed as a second choice, the arm movements would be displayed as separate actions of upper arm, forearm and hand. EWMN defaults to the exact opposite (see Sects. 2.2.1 and 2.2.3–2.2.5). Each of the approaches show certain advantages in different analytical contexts. The benefit for MovEngine lies in the variety of available options and in the ability to choose from the most suitable approaches to represent movement (for details see Sects. 3.2–3.6).

2.4 Additional Concepts

While the notational framework addresses many issues previously neglected in 3D animation, it does not always provide a complete solution for producing realistic movement. While notation systems provide many details of performance, they still rely on interpretation by the reader/dancer, which can greatly reduce the amount of information contained in a score. Currently, MovEngine cannot perform this human-like interpretation. Therefore, in many instances a MovEngine score will contain more movement details than a corresponding notation score. For example, forearm movements frequently invoke implicit rotational movements of the upper arm [21]. In MovEngine, these rotational movements need to be given explicitly. Furthermore, some extra information is needed to determine the adequate acceleration and deceleration within defined movement paths to create realistic movement dynamics (see Sect. 3.2).

In the future some interpretational capabilities might be added to MovEngine, which should simplify the creation of a MovEngine score and make these more similar to traditional notation scores. To accomplish this, reliable algorithms to perform these interpretations need to found, considering anatomical constraints and physical properties of the moving body. The current limitations, however, provide a chance to manually explore and discover many movement details which usually remain on an unconscious level. The ability to manually set and change detailed parameters of movement should be considered an important feature, which should be maintained even if advanced automated procedures were available.

3 MovEngine Design and Language

3.1 User Interface

Currently MovEngine consists of the core animation library and a minimal user interface to create and edit animation scores. Furthermore, necessary tools are provided to visualize and animate the movement phrases. The main application screen consists of two essential view ports: the animation view (Fig. 11), and a score view (Fig. 12) that shows the movement commands on a time line.

After an animation has been loaded from a file, the movement instructions may be seen and edited in a table-like score, which is constructed in a similar fashion to an EWMN staff. The limbs of the animated figure are listed in the left most column of the table. The top row shows time indications which are given in milliseconds. Colored rectangular sections, which are placed in the limb rows, symbolize the times and durations a limb is moving at. Each of these rectangles represents a

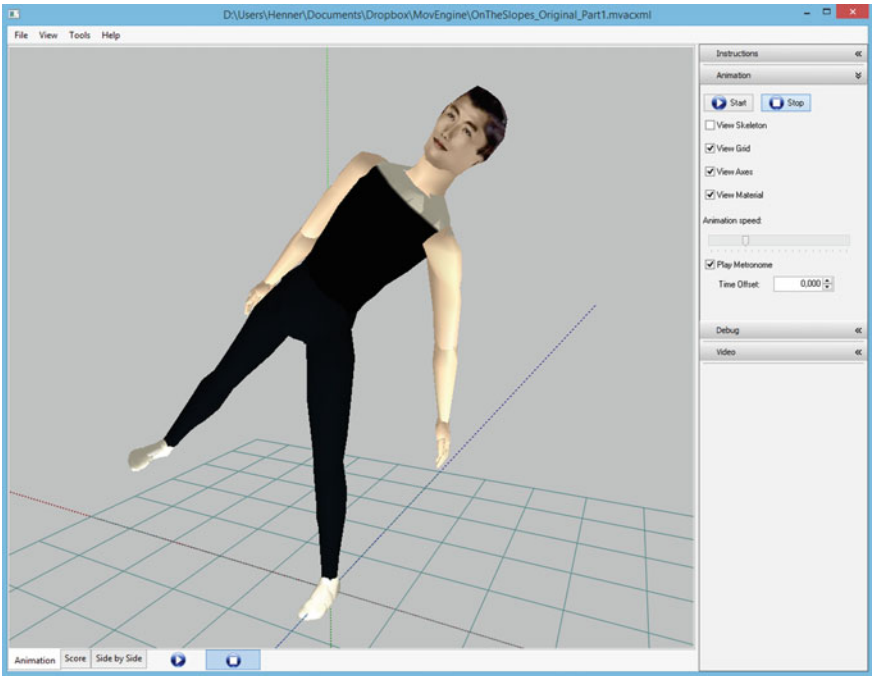


Fig. 11 MovEngine animation

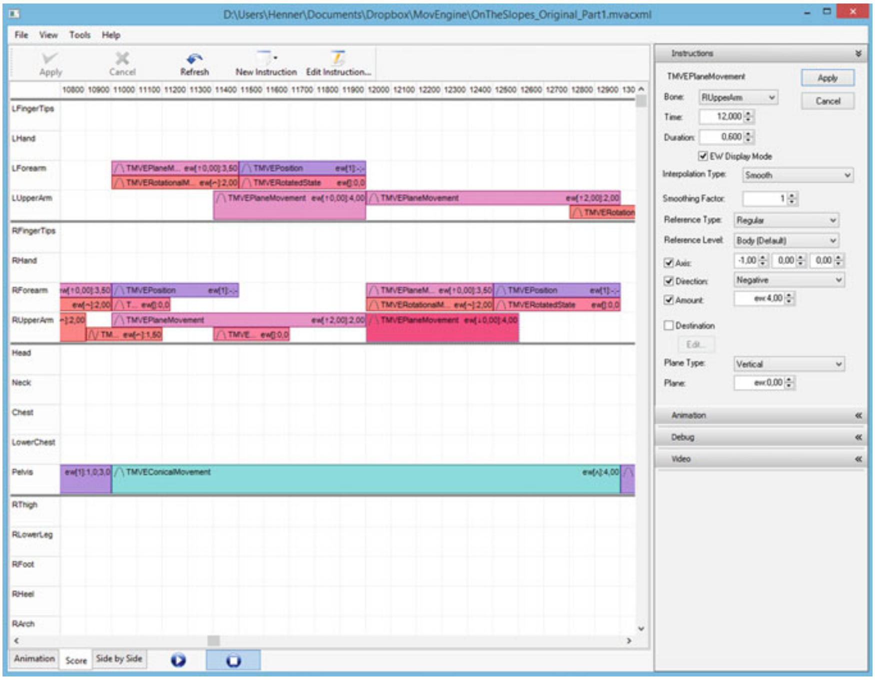


Fig. 12 Time line view of an animation

movement instruction. Different colors are used for marking various types of movement, such as direction changes of a limb (planar or conical movement) or rotations about the axis of a limb. The orchestration of instructions applied to several limbs provides the complex information needed to analyze and describe the movement of the animated figure.

3.2 *Movement Instructions*

The movement instruction approach is directly adopted from the movement notation systems: An instruction is given, when a part of the body changes, and these changes may be described in a variety of ways. One possibility is to state a destination, but a change may be also defined by stating the movement path and magnitude in relation to the starting position. This is in contrast to the key frame approach usually taken in animation software, which focuses on key framed positions. In key frame animation movement is only implicitly encoded as the difference between two frames. The specific ways movement is encoded in the instructions were inspired by the various language elements of EWMN and KIN.

Each movement instruction relies on three parameters: part of the body which is moving, start time and duration. In addition to these parameters, which were adopted from the notation systems, MovEngine requires an additional piece of information to create animations with realistic movement dynamics. Accelerations and decelerations performed at the transitions between the single movements are usually not represented in movement notation. Smooth transitions and the appropriate flow in movement will be created by a human reader, who is able to integrate physical experience into the interpretation of the score.

MovEngine does not have this interpretational ability. Therefore, an additional parameter sets the interpolation algorithm for each movement instruction and thus influences the timing of the movement and the transitions to the preceding and following instructions. For example, a movement may accelerate from or decelerate to zero velocity, or one can create smooth transitions between instructions using spline algorithms.

3.3 EWMN Movement Paths

EWMN provides a highly systematic and concise way to describe the spatial paths taken be the moving limbs as described in Sect. 2.2.3. Because of the accuracy and flexibility of this approach—virtually every possible movement of free limbs and extremities are representable this way—this model was the first to be implemented in MovEngine. The ability to reduce any movement definition down to a rotation around an axis of movement facilitated the implementation of the required algorithms. For each of the EWMN types of movement—plane, cone and rotation—a corresponding instruction class was created in MovEngine.

If a movement is generated by simple interpolation between a start and an end position, like in traditional key frame animation, the resulting direct connection between those two positions will be always a plane and/or a rotational movement. Conical movements are indirect, circular connections between two points, needing additional intermediate points for the interpolation. Also plane movements and rotational movements, which encompass an angle equal or larger than 180°, require additional intermediate key frames to clarify the passage of the limb. As opposed to that, the EWMN approach is far more elegant. Here it is possible to define a path of a full or even longer circle in one single instruction, emphasizing the movement path as opposed to the positional destination.

The approach to focus on single limb movements relies on *Forward Kinematics*. Generally, in 3D character animation *Inverse Kinematics* are considered more suitable for goal-orientated movements and provide immediate solutions in many situations. However, the MovEngine approach—based on forward kinematics and driven by EWMN—provides a high degree of flexibility and accuracy for the exploration of movement style and quality through a descriptive language. Distinct movement properties are exposed on a very detailed level—individually for each moving limb. The focus on single limbs is an important feature, as it is able to keep the core of the movement language concise and simple. This is accomplished by adhering to the model of circular movement: The movement possibilities of a single skeletal limb segment are limited and due to the fixed bone length, the points in reach of the limb always form a sphere.

As opposed to that the resulting spatial paths of several limbs moving in concert are highly complex and it is impossible to describe this complexity directly with a limited set of instructions. In the practical usage of EWMN in movement composition and teaching, the control of movements in individual limbs and their combination and coordination has been applied and explored extensively. MovEngine allows to follow this analytical approach called *simultaneous movement* (see Sect. 2.2.4) and provides the ability to adjust timing of individual movement instructions in adjacent limbs resulting in a high degree of fine control for coordination and for the created movement paths.

3.4 Law of Light and Heavy Limbs

MovEngine respects the EWMN *Law of Light and Heavy Limbs* (see Sect. 2.2.4) and inverts the hierarchy of limbs depending on the supporting body parts. Special contact instructions will establish a firm connection to the ground if a certain part of the body functions as a supporting base. The origin of the hierarchy of limbs will be set at the contact point, and following this hierarchy one can proceed to the freely moving extremities. In stepping and turning the contact point may also act as a virtual joint between the ground and the body, emulating the gravitational influence and the mechanics between the ground and the supporting base. The dynamic inversion of the limb hierarchy also simplifies the modeling of certain movements to a great extent. For example, a pelvic tilt performed on one supporting leg will not affect the supporting leg. Its foot will remain unchanged on the ground. The other, free leg, however, will be carried by the pelvic movement and will be lifted (see Fig. 11).

3.5 Progression and Steps

Although MovEngine substantially relies on EWMN analytical principles, some important procedures borrow ideas from KIN movement analysis. This is especially true in the implementation of progression in space and steps. As mentioned before, this feature is still in the process of being implemented, but the necessary analytical approaches have been fully developed.

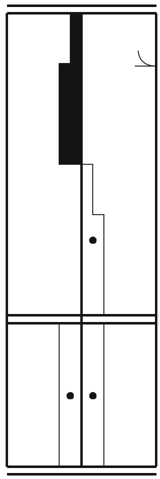
The KIN approach to represent progression and steps analyzes the movement of the center of gravity in relation to the supporting base, which usually changes from on foot to the other. It mainly consists out of two parameters: The direction of the horizontal shift of weight and the distance of the center of gravity to the floor. In KIN terminology the latter is called level. Movements of low level are performed on bent legs, movements of middle level on straight legs. High level movements require the performer to rise on the toes. If the support relies on other parts of the body than the feet, certain conventions define the distance of the center of gravity to the ground.

EWMN's approach may be regarded as similar to this, but it is different in some important details. In stepping the center of gravity does not play the most important role. Instead, the movements of the supporting and gesturing leg are emphasized. Most notably, the equivalent to the KIN level is not displayed as a movement of the center of gravity, but as bending and stretching movements of the supporting leg.

From a computational view, the KIN approach is easier to implement with its focus on the center of gravity. The relation from the contact to the ground towards the moving center of gravity and their connection through the linkage of the supporting leg may be best modeled by inverse kinematics. Likewise, the movement of gesturing leg needs to be resolved by inverse kinematics while the foot is moving in the direction of progression.

With this basic preference to follow the KIN way to model steps, there is still another important property about the analysis in EWMN, which needs to be considered. In KIN in many cases all the necessary instructions to perform a step are embedded in one symbol (Fig. 13), while in EWMN this information is usually distributed over several symbols (Fig. 14). In EWMN release and contact symbols explicitly define the exact timing when a contact is established, and when the foot leaves the floor. In KIN this information is usually given implicitly in the single direction symbol, although there are also possibilities to define this information in more explicit ways. Furthermore, in KIN the direction of the step and the level information is also given in one symbol. For realistic animation results, changes of level might occur at slightly different timings than the horizontal shift of weight. Therefore MovEngine should split this information into different instructions, which can be timed independently. Generally, the EWMN way of explicitly stating

Fig. 13 Example of a simple step sequence in KIN



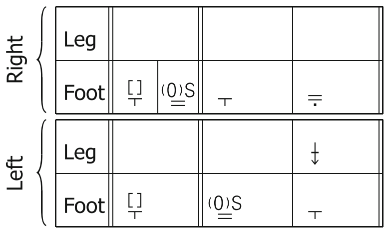


Fig. 14 The same step sequence in EWMN

contact and movement details separately as independent instructions is more suitable for the computations in MovEngine, as no interpretational measures are required. It also provides greater flexibility to create different ways and styles of stepping only by changing the timing of the individual components.

3.6 Low-Level Versus High-Level Description

The notation-based framework acts as a concise low-level description language for the animated dance movement, providing comprehensive motion commands. To avoid constant and repetitive design of movement phrases out of low-level primitives, it is planned to introduce a mechanism to design larger building blocks in future versions. These reusable high-level description entities will play an essential role in user interaction. Once suitably designed, they will free the user of the need to interact with symbolic language alien to the dance style of interest. However, to fine-tune the representation and to adjust nuances of a dance style, the software will always allow for editing of low-level movement commands.

Libraries of high-level language entities will be designed and assembled, to assist in the construction and re-composition of dance material. These language elements may correspond to e.g. choreographic design patterns or identified steps in a certain dance style. They can also relate to more complex concepts introduced by dance notations, like many of the conventionalized expressions used in the KIN system. All library entries will be directly or indirectly linked to low-level commands. Indirect linking will be accomplished by setting up a chain of linked entries, which will ultimately lead to the constituting low-level commands. As far as possible, library entries shall be designed with open parameters, to allow for flexible, ad hoc modifications of a certain movement.

The true strength of this design is revealed, when the constituting lower-level building blocks of library entries are edited, after they have been linked to an animation. All instances of the edited entry will change their appearance accordingly, allowing the researcher to control characteristics of movement transitions and style through a centralized mechanism. Thus an animated choreographic sequence may undergo many experimental variations at only a few mouse-clicks.

A researcher is given as much control as possible to creatively pursue the movement potential hidden in incomplete sketches, movement concepts or historic sources.

4 Current Usage and Outlook

In the current implementation status of MovEngine the core functionality of moving and synchronizing free extremities in space have been implemented and are working as expected. Transitions between the single movement instructions can be influenced by choosing from different interpolation algorithms creating accelerations, decelerations or smooth connections. Floor contacts are being processed, setting and influencing the hierarchy between limb segments. The shifting of weight from one foot to the other and progression in space is currently being implemented, and will be available in the near future.

During the 2012/13 academic year MovEngine has been introduced as a learning tool in movement notation studies at Folkwang University of Arts in Essen. Students are given the opportunity to explore movement phrases by assembling animations out of their notation-based, atomic components. Valuable information is gained in these processes through the visual feedback provided by the animation. Knowledge on movement notation acquired in traditional notation classes may be applied and verified. In addition, two further aspects emerge when working with this visual tool:

First, notation systems as EWMN and KIN are not based on anatomical limitations. They describe an abstract body moving according to spatial and temporal information. The human reader automatically interprets these instructions, considering his anatomical possibilities. Therefore in many instances, notation scores may be simplified and do not need to respect some performance details. MovEngine does not supply this kind of interpretation. Therefore animation scores need to respect many anatomical properties. They need to differ substantially from their notation counterpart to produce a realistic movement appearance. While this requires additional attention by the user, it also provides an excellent chance to teach important anatomical and mechanical facts on human movement, which in regular dance training usually remain on an unconscious level.

The second aspect focuses on the spatial paths drawn by limbs when moving in space. The animated movement output generated by MovEngine is based on the analytic principles of EWMN: Each single skeletal segment can only perform circular movement paths lying on a sphere, as the free end of the limb revolves around its fixed joint. This provides a manageable base for calculating the movement in space. When adjacent limbs move on these circular paths simultaneously, the outer limbs move on their circular path relatively to their fixed joint, but more complex shapes are created in relation to space. MovEngine can draw these movement paths in space as colored traces and thus reveal and emphasize essential

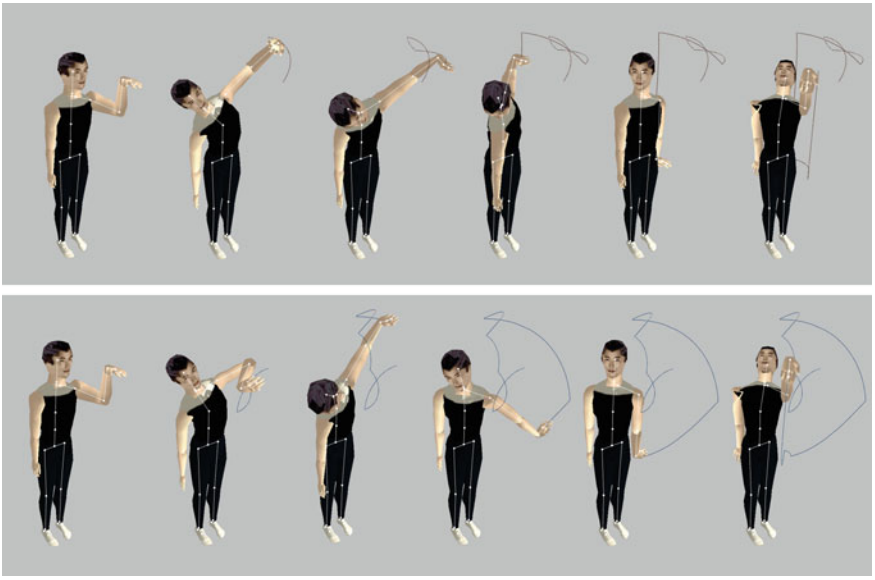


Fig. 15 Examples of a spatial trace of the left hand in a MovEngine-generated animation. The movement instructions of both examples are almost identical. The upper body movement consists of consecutive bending and rotational actions. The order of these two different type of actions differ slightly in the *first* and *second* example. Nevertheless, the resulting spatial paths clearly exhibit different movement qualities. While the first path is almost performed in one vertical plane, the second one extends into various directions creating a more three-dimensional shape. Animation created by Foteini Papadopoulou [22]

details in dance performance, which are hard to grasp without additional aids. It is interesting to note, that minute changes in quantity of the single circular movements or temporal changes in their orchestration, produce enormous changes in the spatial result. Dancers and choreographers get an opportunity to visualize these otherwise hidden properties, and furthermore a chance to study which actions and changes are needed to produce a certain result (see Fig. 15). Interesting enough, these spatial traces were the subject of Eshkol's computer-based research in the late 1960s at the BCL, University of Illinois, which was mentioned in Sect. 1.1. At that time, the computation and visualization of spatial traces were a major endeavor and the sole subject of several years of research. Today, MovEngine uses the same foundation and computations to animate a human-like figure, while the visualization of the space-chords is a more than welcome by-product of larger goals.

While MovEngine is currently being presented as one software package with a distinct set of features, its core actually consists of a multi-purpose library which is responsible for the generation of animated movement out of notation-based instructions. The current application layer only provides tools to test and verify the functionality of the core library in an experimental context. As a long term goal, various specialized applications could be built on top of the MovEngine library benefiting from its notation-based approach but providing different sets of features and substantially different interfaces to the user.

The original research project *Visualizing (the Derra de Moroda) Dance Archives* [1] proposed a reconstruction tool to visualize and animate static material (sketches, verbal descriptions, pictures) on historic dances. Also MovEngine could serve as tool for visual movement composition providing means to create, combine and vary complex movement phrases. Specialized applications could intentionally expose the details of the underlying notation-based movement analysis to the user, or they might choose to hide those details and provide more accessible, visual means to present the movement data. While MovEngine internally processes movement through its native language based on the notation systems EWMN and KIN, the movement content could be translated and communicated to the user using only one of the systems or in any other feasible way.

MovEngine could supply an important contribution to robotic research through its descriptive movement language. In contrast to the original systems of movement notation the MovEngine language is fully suited for machine-driven processing. The data does not need to be interpreted by a human reader, like notation scores do. It describes movement in explicit ways and exposes distinct properties to control the movement on a detailed level. Still, the MovEngine language employs the important notational principles of reducing the complexity of movement data through segmentation and abstraction. However, an application in a robotic context needs to respect the physical properties of mass and anatomical constraints of the robot which are not considered essential in the current context of animation. Therefore, an extra layer will be needed to process these constraints. The inclusion of MovEngine's descriptive language can be especially beneficial, if the focus lies not only on generating goal-orientated movements, but also on creating and carefully adjusting movement paths in space.

As a device that offers immediate visual feedback, MovEngine may contribute to the discussions among researchers, notation experts and practitioners. It can help to articulate the implicit knowledge of movement and in that way contribute to research on dance and to the discussion about movement in general. Nevertheless, the thorough and faithful exploration of well-proven, but not widely known methods of movement description provides the strength of MovEngine's unique approach. The transference and translation of traditional knowledge to the visual realm creates new opportunities to revive notation-based approaches, which otherwise tend to loose their importance in the era of video and digital processing. It is expected that more traditional analytical aspects will emerge in the continuing course of development and application that will serve new purposes and will gain new meaning through the process of visualization.

References

1. H. Drewes, C. Jeschke, Visualizing (the Derra de Moroda) Dance Archives. www.movement-notation.org/VisualizingArchives (2013)
2. N. Eshkol, A. Wachman, *Movement Notation* (Weidenfeld and Nicolson, London, 1958)
3. R. Laban, *Schrift-Tanz* (Universal Edition, Wien, 1928)
4. N. Eshkol, Notation of Movement. BCL report 10. (University of Illinois, Urbana, Illinois, 1970)
5. E. Wilson, Une Espèce de Sympathie. Eskol's Spheres and Lockhart's Loxodromes, in *Sharon Lockhart|Noa Eshkol (exhibition catalogue)* (Sternberg Press, Berlin, 2012), pp. 59–64
6. M. Leeker, EWMN. Toward a cooperation of self-organized actors in digital cultures, in *Sharon Lockhart|Noa Eshkol (exhibition catalogue)* (Sternberg Press, Berlin, 2012), pp. 97–102
7. S.W. Smoliar, L. Weber, Using the computer for a semantic representation of labanotation, in *Computing in the Humanities* (University of Waterloo Press, Waterloo 1977), pp. 253–261
8. D. Herbison-Evans, LED & LINTEL: a windows mini-editor and interpreter for labanotation. http://donhe.topcities.com/pubs/led.html. Revised 15 Nov 2012
9. T. Calvert, L. Wilke, R. Ryman, I. Fox, Application of computers to dance. IEEE Comput. Graph. Appl. **25**(2), 6–12 (2005)
10. Credo Interactive Homepage, Credo Interactive Inc.. http://www.credo-interactive.com (2015)
11. T. Schiphorst, Merce Cunningham: making dances with the computer. Choreography Dance Int. J. **4**(3), 79–97 (1997)
12. P. Teitelbaum, O. Teitelbaum, J. Nye, J. Fryman, R.G. Maurer, Movement analysis in infancy may be useful for early diagnosis of autism. Proc. Nat. Acad. Sci. **95**, 13982–13987 (1998)
13. I. Golani, A mobility gradient in the organization of vertebrate movement: the perception of movement through symbolic language. Behav. Brain Sci. **15**, 249–308 (1992)
14. R. Benesh, J. Benesh, *An Introduction to Benesh Movement Notation* (A. and C. Black, London, 1956)
15. N. Eshkol, *Fifty Lessons by Dr. Moshe Feldenkrais* (Movement Notation Society, Holon, Israel, 1980)
16. N. Eshkol, *Angles and Angels* (Movement Notation Society/Tel Aviv University, Tel Aviv, 1990)
17. J. Harries, T. Sapir, *About Time in Eshkol-Wachman Movement Notation* (Mayshav Publishing, Tel Aviv, 2009)
18. T. Sapir, N. Al-Dor, Moving Landscape. 12 Dances. Composition: Tirza Sapir (Movement Notation Society, Holon 2007)
19. A. Knust, *Dictionary of Kinetography Laban* (MacDonald and Evans, London, 1979)
20. A.H. Guest, *Labanotation—The System of Analyzing and Recording Movement* (Routledge, New York 2005)
21. H. Drewes, L. Renneke, Visualizing (the Derra de Moroda) dance archives. The development of a software tool for dance and movement research, in *Proceedings of the Twenty-Seventh Biennial ICKL Conference* (International Council of Kinetography Laban, Budapest 2012), pp. 183–193
22. F. Papadopoulou, Praxisorientierte Bewegungsanalyse: Prozesse des Komponierens und der Notation am Beispiel der choreographischen Arbeit WörterKörper. MA project (MA Dance Composition—Movement Notation/Movement Analysis) (Folkwang University of the Arts, Essen 2013). (not published)

Bayesian Approaches for Learning of Primitive-Based Compact Representations of Complex Human Activities

Dominik Endres, Enrico Chiovetto and Martin A. Giese

Abstract Human full-body activities, such as choreographed dances, are comprised of sequences of individual actions. Research in motor control shows that such individual actions can be approximated by superpositions of simplified elements, called movement primitives. Such primitives can be employed to model complex coordinated movements, as occurring in martial arts or dance. In this chapter, we will briefly outline several biologically-inspired definitions of movement primitives and will discuss a new algorithm that unifies many existing models and which identifies such primitives with higher accuracy than alternative unsupervised learning techniques. We combine this algorithm with methods from Bayesian inference to optimize the complexity of the learned models and to identify automatically the best generative model underlying the identification of such primitives. We also discuss efficient probabilistic methods for the automatic segmentation of action sequences. The developed unsupervised segmentation method is based on Bayesian binning, an algorithm that models a longer data stream by the concatenation of an optimal number of segments, at the same time estimating the optimal temporal boundaries between those segments. Applying this algorithm to motion capture data from a TaeKwonDo form, and comparing the automatically generated segmentation results with human psychophysical data, we found a good agreement between automatically generated segmentations and human performance. Furthermore, the segments agree with the minimum jerk hypothesis about human movement [32]. These results suggest that a similar approach might be

D. Endres (✉)
Theoretical Neuroscience Group, Department of Psychology,
Philipps-University Marburg, Marburg, Germany
e-mail: dominik.endres@uni-marburg.de

D. Endres · E. Chiovetto · M.A. Giese
Section Computational Sensomotorics, Department Cognitive Neurology,
University Clinic Tübingen, Tübingen, Germany
e-mail: enrico.chiovetto@uni-tuebingen.de

M.A. Giese
e-mail: martin.giese@uni-tuebingen.de

J.-P. Laumond and N. Abe (eds.), *Dance Notations and Robot Motion*,
Springer Tracts in Advanced Robotics 111,
DOI 10.1007/978-3-319-25739-6_6

useful for the decomposition of dances into primitive-like movement components, providing a new approach for the derivation of compressed descriptions of dances that is based on principles from biological motor control.

1 Introduction

Like choreographed dances, complex human full-body activities are comprised of sequences of individual actions. For purposes of teaching and memorization, different often heuristically motivated methods for the abstract notation of such movement sequences have been proposed, e.g. in terms of dance step diagrams or schemes for the execution of forms in martial arts, e.g. katas in Karate, or Hyeongs and Taegueks in TaeKwonDo. Viewed from a computational perspective, these diagrams are compressed versions of the movements, which have to be decompressed by the actors or dancers during execution. This decompression is possible because the dancers are able to 'fill in the blanks' between subsequent foot positions in a dance step diagram using their own motor repertoire.

A variety of techniques for the modeling of complex human behavior sequences has been proposed in computer science, and a review would exceed the scope of this chapter. Instead we present here several examples from our work where we try to exploit biological concepts to derive mathematical models for such complex human behaviors.

According to a prevalent hypothesis in human motor control, complex coordination patterns within individual movements are organized in terms of movement primitives, i.e. simplified control elements, which can be combined in space and time to whole classes of complex movements. In the biological literature a variety of algorithms have been proposed to estimate such primitives from kinematic or EMG data [8, 31].

In Sect. 2.1, we will review several popular definitions of movement primitives (MP). As an example of a state-of-the art algorithm for the unsupervised extraction of MPs from motion capture data, we describe the Fourier-based anechoic demixing (FADA) algorithm [16] in Sect. 2.2 and show that this algorithm outperforms other learning techniques. In addition, we present in Sect. 2.3 a Bayesian approach for the estimation of the model type and optimal number of primitives [24]. This number is also called the *model order*. We demonstrate that our Bayesian approach results in better model type and order estimates than previously applied schemes.

Most MP extraction methods require a previous segmentation of action streams in individual movements that can be characterized by individually controlled actions (periodic or non-periodic). Therefore, an important question is how to determine such segments from longer action sequences. In Sect. 3, we discuss an efficient probabilistic methods for automatic segmentation. The developed unsupervised segmentation method is based on Bayesian binning (BB), an algorithm that models a longer data stream by the concatenation of an optimal number of segments, at the same time estimating the optimal temporal boundaries between those segments. Applying this algorithm to motion capture data from a TaeKwonDo Taeguek, and comparing the automatically generated segmentation results with human psychophysical data, we found a good agreement between automatically generated segmentations and human performance [25, 26]. This was in particular the case when the joint angle trajectories within the segments were modeled by polynomials of order four. This order is consistent with optimal control-based theories of human movements [32] that have been validated in many previous experiments. In particular, these polynomials minimize integrated jerk for given endpoint constraints. Intuitively, this results in movements being as smooth as possible within each segment.

To illustrate that our segmentation approach might be useful for the compressed representation of movements, we create a movement diagram for the *Taeguek Pal-Chang* and compare it to a traditional diagram used in TaeKwonDo teaching (see Fig. 4). Our results suggest that a similar approach could also be useful for the decomposition of dances into primitive-like movement components, serving as a new form of data-driven method for the derivation of compressed dance descriptions.

2 Kinematic Movement Primitives: An Overview

2.1 Definitions of Movement Primitives

A long-standing hypothesis in the neuroscience community is that the central nervous system (CNS) generates complex motor behaviors by combining a small number of stereotyped components, also referred to as muscle synergies or movement primitives [8, 31]. Such components consist of movement variables, such as joint trajectories [44, 67] or muscle activations [14, 18, 19, 42] that are activated synergistically over time.

Different conceptual definitions of movement primitives have been given in the literature, depending on the mathematical models used to factorize kinematic or electromyographic (EMG) data into different types of temporal, spatial, or spatio-temporal components. One classical definition of movement primitive is based on the idea that groups of degrees of freedom (dofs) might show instantaneous covariations, reflecting a coordinated recruitment of multiple muscles or joints. This implies the assumption that the ratios of the signals characterizing the different dofs remain constant over time. This type of movement primitive has been applied in particular in muscle space, where muscle synergies have been defined as weighted groups of muscle activations [12, 72, 73]. Such synergies have also been referred to as “synchronous” synergies, since the different muscles are assumed to be activated synchronously without time delays between different muscles. An alternative way to characterize movement primitives is based on the idea that they express invariance across time, so that they can be expressed as basic temporal patterns, defined by functions of time that are combined or superposed in order to reconstruct the movement signals (EMG signals or joint angles). Temporal components based on this definition have been identified in kinematic [6, 17, 44], dynamic [71] and EMG signal space [13, 42, 43]. “Time-varying synergies” [18–20] have been described instead as spatiotemporal patterns of muscle activation, with the EMG output specified by the amplitude and time lag of the recruitment of each synergy. More recently, also models based on the combinations of other definitions have been proposed. For example, Delis and colleagues [21] defined space-by-time EMG organization patterns, where EMG activation patterns are obtained by mixtures of both temporal and synchronous synergies.

2.2 Unsupervised Learning Techniques for the Identification of Movement Primitives

In the literature, a variety of unsupervised learning methods have been used for the identification of movement primitives from experimental data sets. This includes well-known classical unsupervised learning techniques based on instantaneous mixture models, such as principal component analysis (PCA) and independent component analysis (ICA) [14, 22], and also more advanced techniques that include, for instance, the estimation of temporal delays of the relevant mixture components. An example is the work by d’Avella and colleagues [18, 19], who extended the classic non-negative matrix factorization (NMF) algorithm introduced by Lee and Seug [50] to identify spatiotemporal EMG synergies. Omlor and Giese [58–60] developed a new algorithm based on the Wigner-Ville Transform for the extraction of time-shifted temporal components that is based on an anechoic mixture model (1), as used in acoustics for the modeling of acoustic mixtures in reverberation-free rooms [9, 23, 79]. This model assumes that a set of N_s recorded acoustic signals $x_i, i = 1, 2, \ldots N_s$, is caused by the superposition of N acoustic source functions (signals) $s_j(t)$, where time-shifted versions of these source functions are linearly superposed with the mixing weights a_{ij}. The time shifts are given by the time delays τ_{ij}, and in the acoustical model are determined by the traveling times of the signals. The model has the following mathematical form:

$$x_i(t) = \sum_{j=1}^{N} a_{ij} s_j(t - \tau_{ij}) \tag{1}$$

For the special case that $\tau_{ij} = 0$ for all pairs (i, j), this model (1) coincides with the classical linear combination models underlying PCA and ICA. To analyze kinematic data associated with a pointing task accomplished during locomotion, and inspired by the previous work by Omlor and Giese [58, 59], we [16] developed recently a new algorithm (Fourier-based Anechoic Demixing Algorithm, FADA), that is based on the same generative model (1), but includes additional smoothness priors for the identified functions. The introduction of such priors is justified because EMG or kinematic data from motor tasks usually have limited band-width, and it substantially improves the robustness of the estimation method. Band limited source functions in (1) can be approximated by a truncated Fourier series of the form:

$$x_i(t) \cong \sum_{k=-M}^{M} c_{ik} e^{\frac{2\pi \imath k t}{T}} \tag{2}$$

and

$$s_j(t - \tau_{ij}) \cong \sum_{k=-M}^{M} v_{jk} e^{-\imath k \tau_{ij}} e^{\frac{2\pi \imath k t}{T}} \tag{3}$$

M being a positive integer which is determined by Shannon's theorem according to the limit frequency of the signals. The symbol $\imath$ signifies the imaginary unit, c_{ik} and v_{jk} complex numbers ($c_{ik} = |c_{ik}| e^{\imath \phi_{c_{ik}}}$ and $v_{jk} = |v_{jk}| e^{\imath \phi_{v_{jk}}}$) and T the total number of time samples. Substituting (2) and (3) in (1), and assuming uncorrelatedness of the sources $s_j(t)$ as it was in other previous works [58, 59], the following iterative algorithm can be derived for the identification of the unknown parameters in model (1).

After a random initialization of the estimated parameters, the following steps are carried out until convergence:

1. Compute the absolute values of the coefficients c_{ik} and solve the following positive demixing problem using positive ICA or non-negative matrix factorization:

$$|c_{ik}|^2 = \sum_{j=1}^{N} |a_{ij}|^2 |v_{jk}|^2 \tag{4}$$

   with $i = 0, 1, \ldots N_s$ and $k = 0, 1, \ldots M$. N is the number of sources. Since the signals are real the Fourier coefficients Equations (2) and (3) for positive and negative indices k are complex conjugates of each other. For this reason it is sufficient to solve the demixing problem by considering only the coefficients with indices $k \geq 0$. For the shown implementation we used non-negative independent component analysis [39] for solving the underlying demixing problem with non-negative components.
2. Initialize $\phi_{v_{jk}} = 0$ for all pairs (j, k) and iterate the following steps:

   a. Update the phases of the Fourier coefficients of the sources, which are defined by the identity $\phi_{v_{jk}} = \text{angle}(v_{jk}) = \arctan(\text{Im}(v_{jk})/(\text{Re}(v_{jk}))$ by solving the following non-linear least square problem

$$\min_{\Phi} \|\mathbf{C} - \mathbf{V}\|_F^2 \tag{5}$$

where $(\mathbf{C})_{ik} = c_{ik}$, $(\mathbf{V})_{ik} = \sum_{j=1}^{N} a_{ij} e^{-\imath k \tau_{ij}} |v_{jk}| e^{\imath \phi_{v_{jk}}}$ and $\Phi_{jk} = \phi_{v_{jk}}$. $\|.\|_F$ indicates the Frobenius norm.

b. Assuming that the source functions $s_j(t)$, defined by the parameters v_{jk} are known, the mixing weights a_{ij} and the delays τ_{ij} can be optimized for each signal x_i by minimization of the following cost function:

$$[\hat{\mathbf{a}}, \hat{\mathbf{t}}] = \arg\min_{\mathbf{a},\mathbf{t}} \|x_i(t) - \mathbf{s}(t, \mathbf{t})'\mathbf{a}\|_F^2 \tag{6}$$

Optimization with respect to **a** and **t** is feasible, assuming uncorrelatedness of the sources and independence of the time delays [70]. The column vector **a** concatenates all weights associated with dof i, i.e. $\mathbf{a} = [a_{i1}, \ldots, a_{iN}]'$. The vector function $\mathbf{s}(t, \mathbf{t}) = [s_1(t - \tau_{i1}), \ldots, s_N(t - \tau_{iN})]'$ concatenates source functions associated with dof i, shifted by the associated time delays.

By the approximation of the signals by a truncated Fourier series, compared to more general algorithms, the FADA algorithm has a substantially smaller number of free parameters that have to be estimated. This makes the algorithm for cases where the underlying assumption about the frequency spectrum of the relevant signals are fulfilled more efficient and robust than more general algorithms (such as general anechoic demixing algorithms [58, 59] or classic PCA and ICA applied to temporal data [22, 42]). Since the underlying optimization problems have fewer local minima, convergence of the algorithm is faster and it is less prone to be trapped in irrelevant local minima. We have confirmed these properties in extensive simulations [15] using synthetic ground-truth data that was derived from known generative models.

For example, the FADA algorithm performed better than other methods when identifying anechoic primitives (Fig. 1). In order to evaluate the performance of different algorithms, artificial kinematic data sets were simulated, using the generative model defined by Eq. (1). The source functions were generated by filtering white noise with a Butterworth filter with a cut off frequency that mimicked the spectrum of real kinematic data. The values of the mixing weights and the time delays were also drawn from uniform distributions over fixed intervals. For the simulations the number of sources was set to four, and the number of simulated signals was 250 for each data set. The data sets were also corrupted with signal-dependent noise drawn from a Gaussian distribution of variance $\sigma = |\alpha x_i(t)|$ [37]. The scaling parameter α was adjusted on order to adjust the correlation between the unperturbed data and the perturbed data, realizing the values $1 - R^2$ equal to 0.05, 0.15, 0.25 and 0.35. The similarity (S) between original and identified primitives (source functions) was quantified by computing the maximum of the scalar products between original and recovered primitive (over all possible time delays).

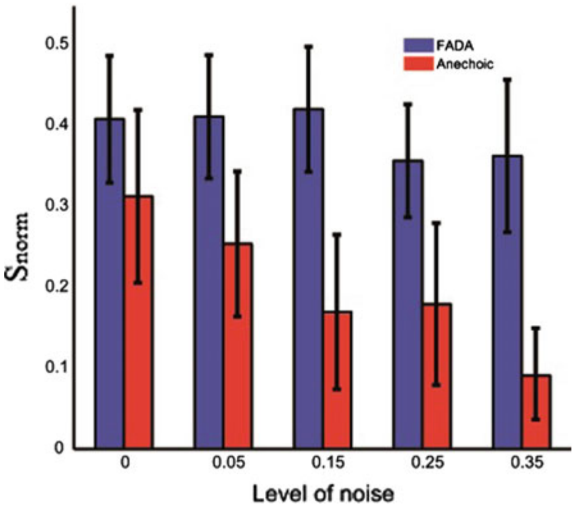


Fig. 1 The figure shows the average level of the similarity between actual and identified anechoic primitives for different levels of signal dependent noise. The primitives were estimated from artificial ground-truth data sets with the anechoic demixing algorithm by Omlor and Giese [58, 59] and the new algorithm FADA [16]. The similarity values are normalized according to the formula $S_{\mathrm{norm}} = (S - S_{\mathrm{b}})/(1 - S_{\mathrm{b}})$, where S_{b} indicates the baseline value that was obtained by assessing the average similarity between the randomly generated source functions. The value $S_{\mathrm{norm}} = 0$ corresponds to chance level similarity, and the maximum value of the normalized similarity is one

2.3 Model Selection Criteria

Different established approaches for the extraction of movement primitives from trajectory and EMG data differ, on the one hand, by the type of generative model that is used (e.g. instantaneous mixtures vs. models containing time delays). On the other hand they also can differ in terms of the number of model parameters, e.g. the number of primitives or source functions in the mixture model. The number of primitives is also called the *model order*. To our knowledge only very few motor control studies have so far addressed the problem of model selection in a principled way, see e.g. Delis and colleagues [21] and Hart and Giszter [38] for notable exceptions. The existing generative models for the extraction of movement primitives have indeed been demonstrated to provide a low-dimensional decomposition of the experimental data, but no clear criterion has been developed to objectively determine which model is best suited for describing the statistical properties of the data under investigation.

In the field of machine learning various methods for the optimization of model complexity have been developed, either using heuristic approaches or methods derived from Bayesian inference. The well-known Akaike Information criterion (AIC) and Bayesian Information Criterion (BIC) have the advantage of being easy to use when a likelihood function for a given model is available. Hence, they are often the first choice for model order estimation, but not necessarily the best one. In the work by Tu and Xu [74] several criteria for probabilistic PCA (or factor analysis) models were evaluated, including AIC, BIC, MIBS [54] (Minka's Bayesian model selection) and Bayesian Ying-Yang [78]. The authors found that MIBS and Bayesian Ying-Yang work best. AIC and BIC criterion have also been used to estimate the number of independent components in fMRI data. This was done for instance by Li and colleagues [52] that, however, found AIC and BIC estimation performance to be adversely dependent on temporal correlations between signals. Other heuristic methods have been used on the literature for model order selection. Such approaches typically utilize some features of the reconstruction error (or conversely, of the variance-accounted-for (VAF)) as a function of the model order. For instance, the usual procedure is to search for a "knee" in that function, a procedure which is inspired by the scree test for factor analysis [11]. For example, multiple authors [12, 18, 42] used them to determine the number of EMG synergies underlying different human behaviors.

To improve the accuracy provided by standard Bayesian estimator, we developed a new objective criterion (which we called *LAP*) in the framework of Bayesian generative model comparison [7] for model-order selection that extends the other classical ones based on information-theoretic and statistical approaches. The criterion is based on a Laplace approximation of the posterior distribution of the parameters of a given blind source separation method, re-formulated in the framework of Bayesian generative model comparison [7]. Exploiting the Laplace approximation allowed us to approximate some intractable integrals appearing in the computation of the marginal likelihood of the models that, after the approximation, assumed the following form:

$$
\begin{aligned}
p(D|\Phi, M) \approx & \underbrace{\log(p(D|\Theta^{\star}, \Phi, M))}_{\log-\text{likelihood}} + \underbrace{\log(p(\Theta^{\star}|\Phi, M))}_{\log-\text{prior}} \\
& + \underbrace{\frac{\dim(\Theta)}{2}\log(2\pi) - \frac{1}{2}\log(|\mathbf{H}|)}_{\log-\text{posterior}-\text{volume}}
\end{aligned}
\tag{7}
$$

where D indicates the observable data, Θ_M is a tuple of model parameters for a model indexed by M (the 'model index') and Φ indicates a tuple of hyperparameters. In addition, $\Theta^{\star}$ is a tuple of model parameters that maximize the log-likelihood subject to the regularization provided by the parameter prior and $\mathbf{H}$ is the Hessian matrix (second derivatives of the log-posterior = log-likelihood + log-prior at $\Theta^{\star}$) [24]. Equation (7) comprises three parts, which can be interpreted. The first term the log-likelihood measures the goodness of fit, similar to explained variance or VAF. The second term is the logarithm of the prior, which corresponds to a regularization term for dealing with under-constrained solutions for Θ when the data set is small. Finally, the third part measures the volume of the parameter posterior, since $\mathbf{H}$ is the posterior precision matrix (inverse covariance) of the parameters in the vicinity of $\Theta^{\star}$, i.e. it indicates how well the data constrain the parameters. Given different models, to discriminate the most suitable one to describe the available data the criterion requires to compute the values of the model evidence (7) associated with each model and to choose the one which maximizes this evidence as the most appropriate model. With a similar procedure, it is also possible to identify, given a specific model, the right model order. This can be done by computing the model evidences associated with different model orders and by taking the order associated with the highest value as the best one to represent the data.

We showed in our previous work [24] how the *LAP* is more reliable than other already existing classical criteria in selecting the generative model underlying a given data set, as well as in determining the best model order. The criterion performance was evaluated on synthesized data and compared to the performance provided by AIC and BIC. Some results adapted from the work of Endres and colleagues [24] are reported in Fig. 2. Panel a shows that the generating model, here an anechoic mixture, is correctly identified by LAP with near certainty for all three testes noise levels. In panel b, we plotted the difference ΔI between estimated and true model order. LAP provides the best estimates across noise levels, if the correct model type (AMM) is used for the estimation procedure.

3 Generating Compressed Movement Descriptions with Bayesian Binning

3.1 Objective and Related Approaches

We now turn to the problem of extracting individual actions from longer streams of activities. This problem is particularly important in dance, where a dance is typically choreographed by concatenating a sequence of individual dance movements. In addition, the approaches described in the last section typically assume previous segmentation into elementary actions, which then can be modeled by superposition of a set of control primitives. This raises the problem of an automatic segmentation of natural action streams, which has been a central problem in movement analysis and modeling for several decades.

One of the earliest approaches achieved automatic segmentation into piecewise linear functions [5], employing dynamic programming for the solution of the resulting least-squares trajectory fitting problem. There, it is also mentioned that the approach might be extended to arbitrary polynomial orders, but exactly how is not shown. An online version of this algorithm for the piecewise linear model was later developed [45], it was subsequently extended to work with polynomials of higher order [51]. The intended application domain is data mining, where online-capable algorithms are required, since the databases are typically too large to fit into working memory. Recently, real-time versions of these algorithms were derived which achieve the necessary speed-up by using orthogonal polynomials as basis functions for segment contents [33, 34].

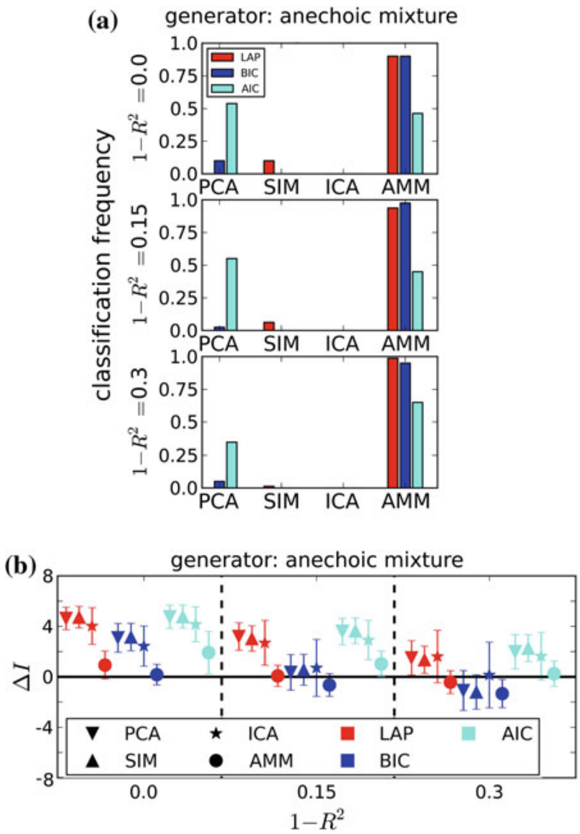


Fig. 2 *Panel* **a** shows the model classification performance provided by different model selection criteria (AIC, BIC and LAP) when applied to artificial data sets (based on model (1) and generated as explained in the previous subsection). The data is corrupted by three different levels of noise. Four different models were tested for each data set, with PCA indicating the probabilistic Principal Component Analysis linear model, Smooth Instantaneous Model (SIM), Independent Component Analysis (ICA) model and Anechoic Mixture Model (AMM). The model selection criteria BIC and LAP both performed well. The AIC criterion frequently confuses PCA and AMM. In panel **b** the estimated number of sources provided by the criteria are displayed. ΔI indicates the difference between the estimated and actual number of primitives. *Symbol shapes* stand for analysis algorithm, *colors* indicate the selection criteria. If AMM was used for analysis, BIC and LAP provided similar estimations for the number of primitives. AIC tended to overestimate the number of primitives. For incorrect analysis model, all criteria provided a higher number of sources. Figure adapted from [24] subject to the creative commons attribution license, version 4.0 (http://creativecommons.org/licenses/by/4.0/)

Most of these approaches use heuristics for the determination of the polynomial order (or leave it up to the user to choose one). We developed an exact Bayesian approach addressing the model complexity issue [25]. One might wonder why polynomials are popular segment models, as opposed to the other movement primitive models described above. The reason for this is twofold: first, polynomials are mathematically convenient and well understood. Second, trajectories that minimize integrated jerk are polynomials of (at most) order five [32], and human movement production seems to implement this optimality principle in many instances. Thus, B-splines have found applications in humanoid robotics, for example for human-robot transfer [75] and goal-directed movement synthesis [76].

Important non-polynomial action segmentation methods are based on hidden Markov models (HMM) [36, 48] (see also Liu et al.'s contribution [53] in this book), change-of-contact events [77], Gaussian mixtures [3], and probabilistic principal component analysis (pPCA) approaches [3]. Because pPCA encodes a full covariance model, it has the potential advantage of handling correlated sensor noise better than models with a polynomial time dependence of the mean and an isotropic noise assumption. We showed [25] how these two approaches can be combined.

However, movement primitive extraction is only one application where action segmentation is interesting: many methods for movement synthesis in computer graphics generate individual segments by (parametric) interpolation between example trajectories, e.g. by PCA [41, 66] or the classical verb-adverbs approach [64], which uses B-splines. For these methods to work, accurately pre-segmented data are required. Similarly, motion graphs [46, 65] or motion database sequencing methods [2] give most convincing results when segment boundaries are consistent with human perception.

Finally, in Psychology and Neuroscience, models for human segmentation performance are expected to provide information about the structure of action representation in the brain, since human brain activity appears to be time-locked to perceived event boundaries [80]. Thus, the segmentation of sequences of piecewise linear movements in the two-dimensional plane was studied [1, 69]. Polyakov et al. [61] fitted monkey scribbling trajectories with parabolic segments, which minimize jerk [32] and comply with the two-thirds power law [49]. They determined that neural signal changes correlate with the segment boundaries thus computed. We compared 3D action segmentation by human observers to polynomial segments computed with Bayesian Binning (BB) [25], finding a good correspondence between 4th-order polynomials (which minimize jerk) and human observers.

In the next section, we describe the basic idea behind BB and apply it to the segmentation of a TaeKwonDo Taeguek. A Taeguek is a stylized martial arts action sequence, in which the performer fights against a number of imaginary opponents. For teaching and memorization purposes, these Taeguek can be described by diagrams of key poses[1] similar to the dance steps diagrams used e.g. in ballroom dancing [10]. This kind of description can be 'decompressed' by a martial artist if he has learned the individual techniques that correspond to the key poses. Our goal is to automatically extract a set of key poses that succinctly describe the Taeguek with BB.

[1]The diagrams for all Taegueks can be viewed on www.taekwondo.de.

3.2 Bayesian Binning

Bayesian Binning (BB) is a method for modeling data by piecewise defined functions. It can be applied if the data have a totally ordered structure, such as a time series. Since it is an (exact) Bayesian approach, it allows for an automatic control of model complexity. In this context, this means that the number of segments (bins), their length and the model for segment contents (the movement primitives) are determined with minimal user intervention. Originally, BB was developed for probability density estimation of neural recordings and their information theoretic evaluation [27]. Later, it was extended for regression of piecewise constant functions [40] and further applications in neuroscience [28, 29]. A similar Bayesian formalism for dealing with multiple change point problems was concurrently developed [30].

In the following subsection, we describe BB in terms of a probabilistic graphical model. For a mathematical treatment of the polynomial movement primitive model, we refer the reader to a previous publication [25]. We also forgo developing the algorithm in detail, since we did that elsewhere [27]. BB operates in discrete time, hence the time axis (see Fig. 3b) is discretized into intervals of duration Δ. Δ has to be small enough so that all relevant features of the data do not change noticeably within such an interval. A natural lower bound on Δ is given by the time resolution of the data recording equipment. The VICON motion capture system which was used for the TaeKwonDo recordings had a frame-rate of 120 Hz, hence $\Delta \geq \frac{1}{120\,\mathrm{Hz}} \approx 8.3\,\mathrm{ms}$. The intervals are labeled by a discrete time index t, which runs from 0 to T. BB concatenates these intervals into contiguous, non-overlapping segments. In the example shown in Fig. 3c, there are two segments: the first one extends from $t = 0$ to $t = i$ (inclusive), the second one comprises time steps $t = i + 1$ to $t = T - 1$. In Fig. 3b, the segment boundaries are indicated by the orange lines (there is an implicit boundary at the end, $t = T$).

Suppose we measured a time series of data points D_t consisting of three-dimensional joint angles for Q joints, i.e. $D_t = (x_t^0, \ldots, x_t^{3Q})$. BB makes two central modeling assumptions about such data: (1) within a segment, the parameters Θ_i of the data generating model do not change, and (2) these models are independent across segments. Hence, the joint probability (density) of the data $(D_i, \ldots, D_j) = D_{i:j}$, that is $P(D_{0:T-1}|\Theta_{0:T-1})$ factorizes across segments, in our example:

$$P(D_{0:T-1}|\Theta_{0:T-1}) = P(D_{0:i}|\Theta_0)P(D_{i+1:T-1}|\Theta_{i+1}). \tag{8}$$

This factorization property, combined with the total order of time points, facilitates efficient evaluation of expectations of segmentation point locations and segment parameters. To understand why, consider the graphical model of BB in Fig. 3a, where we use standard graphical modeling notation [7]: circles are random variables, lines with arrows denote conditional dependencies, and dashed rectangles are gates [55], each of which being controlled by a binary gating variable b_t with a

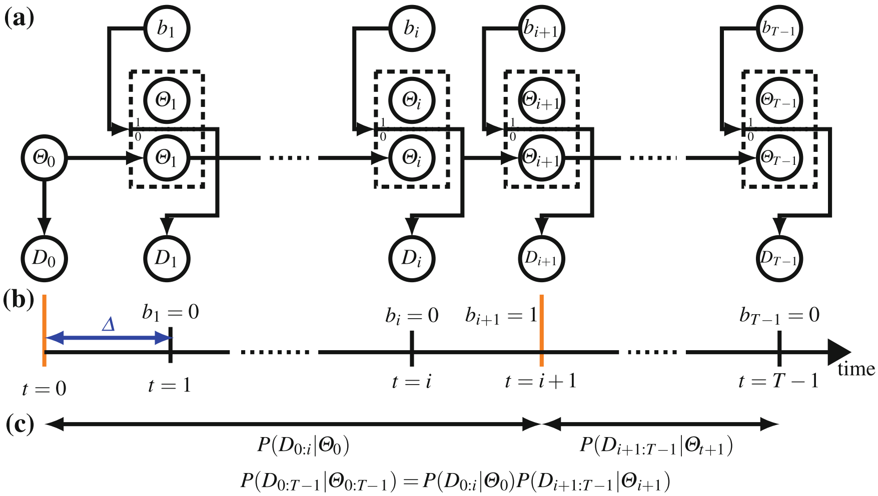


Fig. 3 **a** The graphical model of Bayesian Binning. We follow standard graphical modeling terminology [7]: *circles* represent random variables, *arrows* denote conditional dependencies, and *dashed boxes* are gates [55]. The observable joint angle data D_i are modeled by a (latent) movement primitive model with parameters Θ_i. *Subscripts* denote discrete time steps. Presence or absence of a segment boundary at a time step i is indicated by a binary gating variable b_i: if $b_i = 1$, then the corresponding gate below it is active, which means that a new set of parameters Θ_i is instantiated at this time step. Otherwise, if $b_i = 0$, Θ_i is a copy of the parameters of the previous time step. The graph is singly connected, hence marginal expectations can be computed efficiently with sum-product message passing [47]. For details, see text. **b** The time axis is discretized into non-overlapping, contiguous intervals of duration Δ small enough so that all relevant features of the data are preserved. These intervals are labeled with an integer time index t. There are T such intervals, hence $t \in \{0, \ldots, T-1\}$. Segment boundaries are effectively controlled by the gating variables b_t: in this example, the interval from 0 to $T\Delta$ is subdivided into two segments, as indicated by the *orange* segment boundaries at t = and $t = i + 1$. **c** BB models the time series by contiguous, non-overlapping segments, which are assumed to be independent. With the gating variable setting shown in panel **a** (i.e. only $b_{i+1} = 1$, all others are 0), the joint probability of the data given the parameters $P(D_{0:T-1}|\Theta_{0:T-1})$ factorizes into two contributions: $P(D_{0:i}|\Theta_0)$ for time steps $0, \ldots, i$ and $P(D_{i+1:T-1}|\Theta_{i+1})$ for time steps $i + t, \ldots, T-1$

Bernoulli prior.[2] Depending on the value of this variable, either one or the other alternative part of the model is instantiated. Here, if $b_t = 0$, then the parameters Θ_t for time step t are simply copied from the previous time-step. In our example, this is the case for all $t = 1, \ldots, i$. On the other hand, if $b_t = 1$, the corresponding Θ_t is drawn from a suitably chosen prior distribution (e.g. at $t = i + 1$). This algorithm for determining the parameters effectively concatenates time steps into contiguous, non-overlapping segments. Note that the graphical model is *singly connected*: there

[2] The a priori independent gating variables and their Bernoulli priors induce a Binomial prior on the number of segments, which is a special case of the general priors on segment number boundaries which we developed previously [27]. The latter need a dependency model between the gating variables, which we do not consider here for the sake of simplicity.

is at most one path between any two circles, if one travels along the lines. Hence, the sum-product algorithm [47] can be applied for the efficient and exact evaluation of expectations of interest, if conjugate priors are chosen for the (continuous) parameters.[3] Readers with a machine learning background may observe the similarity of Fig. 3a to a HMM [4, 62], which is probably the most well-known example of a singly-connected graphical model and allows for efficient computation of marginal expectations for the same reason.

The observation model for each time step, given by Eqs. (9)–(11), is a Q-dimensional multivariate Gaussian. It is defined by the parameters $\Theta_i = (z_i, \theta_i^{0,0} \ldots, \theta^{3Q,S})$, that encode a S-th order polynomial time dependence of the mean and a $(3Q \times 3Q)$ covariance matrix Σ_i. The variable z_i specifies the initial time step of the current segment:

$$\mu_i^q = \sum_{s=0}^{S} \theta_i^{q,s} (i - z_i)^s \tag{9}$$

$$\mu_i = (\mu_i^0, \ldots, \mu_i^{3Q}) \tag{10}$$

$$P(D_i|\Theta_i) = \mathcal{N}(\mu_i, \Sigma_i) \tag{11}$$

We showed [25] how to construct a conjugate Gauss-Wishart prior to this observation model (Eqs. (9)–(11)), enabling sum-product message passing.

3.3 TaeKwonDo Taeguek Segmentation

3.3.1 Data Recording

The action streams we studied are TaeKwonDo Taegueks carried out by internationally successful martial artists. Each artist performed the same fixed sequence of 27 kicks and punches of the Taeguek Pal-Chang. The complete form has a full length of about 40 s. We obtained kinematic data by motion capture using a VICON 612 system with 11 cameras. This setup yields 3D positions of 41 passively reflecting markers attached to the performers' joints and limbs with a 3D reconstruction error below 1 mm and at a sampling frequency of 120 Hz.

We used the kinematic data for two purposes. First, we computed joint angles trajectories from a hierarchical kinematic body model (skeleton) which we fitted to the original 3D marker positions. This yielded Euler angles describing the rotations between adjacent skeleton segments in terms of flexion, abduction and rotations

[3] Strictly speaking, any priors that allow for an evaluation of posterior expectations in closed form are suitable, but conjugate priors are particularly convenient.

about the connecting joint (e.g. [57, 63]). Second, from the derived joint angle trajectories we created movie clips showing computer animations of the TaeKwonDo movements. Those videos served as stimuli in a psychophysical experiment to obtain segmentation boundaries according to human perception. For this experiment, we split the motion capture data of each Taeguek manually into sub-sequences of comparable length each containing between three and eight separable TaeKwonDo moves. We presented the videos of these sub-sequences to naïve human observers and asked them to segment the videos by pressing a key. We computed a segmentation density from the key-press timings pooled across all observers, and identified peaks in this density. We refer to the peak locations as ‘population-averaged human segment boundaries’. For details of the experiment and evaluation procedures, see [25, 26].

3.3.2 Segmentation Analysis

We analyzed the joint angle trajectories with BB, as described above. To determine the polynomial order which best matches human perception, we conducted a hit rate analysis [25]: to determine ‘hits’, we counted how often a BB segment boundary was placed in close temporal proximity (±250 ms) to a population-averaged human segment boundary. All BB segment boundaries that could not be matched to a human one were counted as ‘false positives’. The best compromise between hits and false positives is reached at order 4, which agrees with the minimum jerk hypothesis for human movement generation [32]. Put differently, our results indicate that minimum jerk is not only a movement production principle, but also perceptually relevant. Furthermore, the hit rate analysis also revealed that naïve observers’ behavior is best explained by the segment boundaries computed from shoulder and elbow joints taken together. If we were to repeat this analysis with skilled TaeKwonDo martial artists as observers, we would expect to find that the leg joints need to be included in the analysis as well, since kicks are techniques of significant importance in this martial art. This fact, however, was not picked up by the naïve observers in our experiment.

To generate a compressed representation of the Taeguek, we compute the performer’s pose at each segment boundary. The results, for the first part of the Taeguek, are shown in Fig. 4. For visual comparison with TaeKwonDo teaching diagram (Fig. 4, left), we arranged the poses in the same fashion on the right of this figure. Poses match, except for the kicks (pose 2 on the left), pose 2, 3 on the left and the transition (pose 3,1) on the right. This indicates that the compressed description generated by BB may be useful for semi-automatic movement description. We attribute the missing kicks to the fact that we segmented using only arm joints. The naïve observers we tested did not segment at the kicks, too.

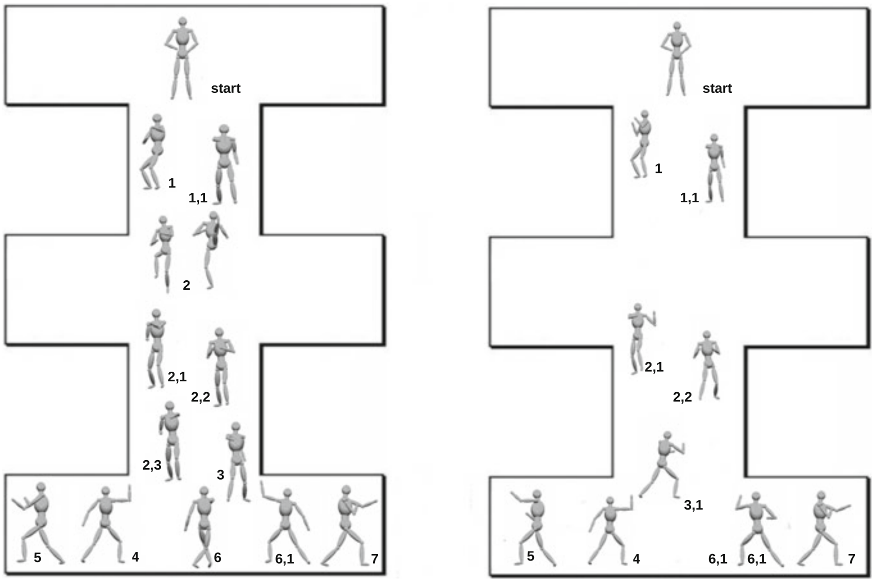


Fig. 4 *Left* Stylized representation of the first part of the movements that comprise the *Taeguek Pal-Chang*, a solo TaeKwonDo form. Only key poses are shown, a trained martial artist can fill in the movements in between these poses using their own motor repertoire. *The bounding box* represents the space within which the martial artist is supposed to move during the Taeguek. Pose numbering corresponds to the numbering found in the diagrams on www.taekwondo.de. *Right* Key poses determined with Bayesian binning employing a 4th order polynomial observation model. This observation model minimizes jerk, and provides the best match to segmentation points determined by naïve human observers, if elbow and shoulder joints are used for the segmentation. Numbering of poses is the same as in the *left panel*. Pose *2,3* is missing here, as are the kicks (number *2* on the *left*), which could not be decoded from arm movements. Pose *3,1*, which is a transition between poses *3* and *4*, does not appear in the Taeguek image. All other poses match

4 Discussion

In this chapter we have summarized some of our recent work that applies biologically-inspired learning-based models for the modeling of complex human activities. At the level of individual actions, we proposed a new efficient algorithm for anechoic demixing that outperforms other approaches for the learning of movement primitives from trajectory and EMG data. In addition we presented how Bayesian inference can be exploited for the selection of the correct generative model and specifically the order of the model. We have shown elsewhere that movement primitives based on anechoic demixing are also suitable for the online synthesis of movements and the embedding in control architectures. For this purpose the learned source function are mapped onto dynamic primitives that are formulated as dynamical systems [35, 56]. At the level of sequences of actions within complex activities, we proposed a new probabilistic method for segmentation that is based on Bayesian Binning, and which applied exact Bayesian inference for determining an optimal segmentation of the action stream into subsequent component actions.

We described an anechoic demixing algorithm in Sect. 2.1. It subsumes as special cases multiple other existing methods for the estimation of movement primitives and can be combined with further constraints on weights and sources, e.g. positivity. In this way, the new model provides a unifying framework for the study of the relationship between different mathematical models for the extraction of movement primitives from data [15]. The investigation of movement primitives, extracted as components of kinematic (joint angle) data seems interesting for dance, on the one hand to characterize the full-body coordination patterns, e.g. between locomotion and upper body movements in dance. On the other hand, such primitives might also be helpful to understand ‘co-articulations’, that is overlaps between subsequent coordination patterns that are important form the smooth flow of motion in dance.

With respect to the method for the segmentation of action streams exploiting Bayesian binning in Sect. 3, our work shows that the obtained segmentation is close to the one provided by humans. However, for other types of movements likely additional features beyond the ones characterizing the arm movements would have to be included. Determining optimal sets of such features from data could also be realized by Bayesian model comparison.

It would also be interesting to investigate if the actor’s poses at the segmentation boundaries are a sufficiently rich description of the movement so that human observers can ‘fill in blanks’ between two subsequent boundaries using their own motor repertoire. In that case, Bayesian binning could be used to derive compressed movement representations similar to dance step diagrams.

As a next step, we could then investigate in how far the obtained compressed representation is suitable for movement production with modular movement primitives. For this purpose, the general polynomial models within the individual segments would have to be replaced by models that are based on primitives of the type discussed in the first part of this chapter.

The modeling assumptions made by BB presuppose that human actions can be viewed as a sequence of clearly separable primitives in time, see Eq. (8). While the match between 4th order polynomial BB and naïve human perception indicate that this presupposition is at least approximately fulfilled for a TaeKwonDo Taeguek, it seems likely that this assumption is violated for other types of human action, such as dancing. There, the transitions from one movement to the next are typically more continuous than in TaeKwonDo. Between-segment dependencies would also be relevant in sign language production, where co-articulation between neighboring signs has been observed [68]. The BB approach can be extended to include dependencies between segments, as long as the graphical model (see Fig. 3a) remains singly connected. Exact inference will then still be possible, albeit with a higher computational effort. Another way of extending our approach would be to include task or context information into the segmentation process [77], to supplement the purely kinematic information which we currently use. Humans use such context information for segmentation when available [81], and rely increasingly on kinematics when context is reduced. Here, again the advantage of the proposed probabilistic approach is that it can be connected easily to probabilistic representations of context and even to semantic representations, as long as they can be expressed in the language of graphical models.

Interesting for online synthesis is also the question in how far individual movements within individual segments can be derived from optimal control problems with boundary conditions derived from the task, refining the rather unspecific polynomial model used for the signal time-courses in the presented model. Ongoing work focuses on this question, aiming at finding optimized segmentations for specific pre-defined control architectures.

Acknowledgments The research leading to these results has received funding from the European Union under grant agreements Koroibot FP7-ICT-2013-10/611909, AMARSi- EC FP7-ICT-248311; FP7-PEOPLE-2011-ITN (Marie Curie): ABC PITN-GA-011-290011, 604102 (HBP), CogIMon H2020 ICT-23-2014/644727, and form the DFG under grants GI 305/4-1, DFG GZ: KA 1258/15-1, and from BMBF grant FKZ: 01GQ1002A. DE has received support from the DFG under grant IRTG-GRK 1901 "The Brain in Action".

References

1. Y. Agam, R. Sekuler, Geometric structure and chunking in reproduction of motion sequences. Journal of Vision **8**(1), 1–12 (2008)
2. O. Arikan, D.A. Forsyth, Interactive motion generation from examples. ACM Trans. Graph. **21**, 483–490 (2002)
3. J. Barbiŏ, A. Safonova, J.-Y. Pan, C. Faloutsos, J.K. Hodgins, N.S. Pollard, Segmenting motion capture data into distinct behaviors, in *Proceedings of Graphics Interface 2004*, GI '04, pp. 185–194, School of Computer Science, University of Waterloo, Waterloo, Ontario, Canada, 2004. Canadian Human-Computer Communications Society
4. L. Baum, T. Petrie, G. Soules, N. Weiss, A maximization technique occuring in the statistical analysis of probabilistic functions of markov chains. Ann. Math. Stat. **41**(1), 164–171
5. R. Bellman, On the approximation of curves by line segments using dynamic programming. Commun. ACM. **4**(6), 284 (1961)
6. B. Berret, F. Bonnetblanc, C. Papaxanthis, T. Pozzo, Modular control of pointing beyond arm's length. J. Neurosci. **29**(1), 191–205 (2009)
7. C.M. Bishop, *Pattern Recognition and Machine Learning*. Springer, Berlin (2007)
8. E. Bizzi, V.C.K. Cheung, A. d'Avella, P. Saltiel, M. Tresch, Combining modules for movement. Brain Res. Rev. **57**(1), 125–133 (2008)
9. P. Bofill, Underdetermined blind separation of delayed sound sources in the frequency domain. Neurocomputing **55**(34), 627–641 (2003). (Evolving Solution with Neural Networks)
10. P. Bottomer, *Ballroom Dancing Step-By-Step*. Anness Publishing, London, UK (2012)
11. R.B. Cattell, The scree test for the number of factors. Multivar. Behav. Res. **1**(2), 245–276 (1966)
12. V.C.K. Cheung, A. d'Avella, M.C. Tresch, E. Bizzi, Central and sensory contributions to the activation and organization of muscle synergies during natural motor behaviors. J. Neurosci. **25**(27), 6419–6434 (2005)
13. E. Chiovetto, B. Berret, I. Delis, S. Panzeri, T. Pozzo, Investigating reduction of dimensionality during single-joint elbow movements: a case study on muscle synergies. Front. Comput. Neurosci. **7**, 11 (2013)

14. E. Chiovetto, B. Berret, T. Pozzo, Tri-dimensional and triphasic muscle organization of whole-body pointing movements. Neuroscience **170**(4), 1223–1238 (2010)
15. E. Chiovetto, A. d' Avella, M.A. Giese, A unifying algorithm for the identification of kinematic and electromyographic motor primitives. Presented at the international conference on the neural control of movement, Puerto Rico, (April 2013)
16. E. Chiovetto, M.A. Giese, Kinematics of the coordination of pointing during locomotion. PLoS ONE **8**(11), e79555 (2013)
17. E. Chiovetto, L. Patan, T. Pozzo, Variant and invariant features characterizing natural and reverse whole-body pointing movements. Exp. Brain Res. **218**(3), 419–431 (2012)
18. A. d'Avella, A. Portone, L. Fernandez, F. Lacquaniti, Control of fast-reaching movements by muscle synergy combinations. J. Neurosci. **26**(30), 7791–7810 (2006)
19. A. d'Avella, P. Saltiel, E. Bizzi, Combinations of muscle synergies in the construction of a natural motor behavior. Nat. Neurosci. **6**(3), 300–308 (2003)
20. A. d'Avella, M.C. Tresch, Modularity in the motor system: decomposition of muscle patterns as combinations of time-varying synergies, in *Advances in Neural Information Processing Systems*, vol. 14, ed. by S.A.S. Michael, I. Jordan, Michael J. Kearns (MIT Press, Cambridge, MA, 2002), pp. 141–148
21. I. Delis, S. Panzeri, T. Pozzo, B. Berret, A unifying model of concurrent spatial and temporal modularity in muscle activity. J. Neurophysiol. **111**(3), 675–693 (2014)
22. N. Dominici, Y.P. Ivanenko, G. Cappellini, A. d'Avella, V. Mond, M. Cicchese, A. Fabiano, T. Silei, A. Di Paolo, C. Giannini, R.E. Poppele, F. Lacquaniti, Locomotor primitives in newborn babies and their development. Science **334**(6058), 997–999 (2011)
23. B. Emile, P. Common, Estimation of time delays between unknown colored signals. Sig. Process. **68**(1), 93–100 (1998)
24. D. Endres, E. Chiovetto, M. Giese, Model selection for the extraction of movement primitives. Front. Comput. Neurosci. **7**, 185 (2013)
25. D. Endres, A. Christensen, L. Omlor, M.A. Giese, Emulating human observers with Bayesian binning: segmentation of action streams. ACM Trans. Appl. Percept. (TAP), **8**(3), 16, 1–12 (2011)
26. D. Endres, A. Christensen, L. Omlor, M.A. Giese, Segmentation of action streams: human observers vs. Bayesian binning, in *KI 2011, LNAI 7006*, ed. by S. Edelkamp, J. Bach (Springer, Berlin, 2011), pp. 75–86
27. D. Endres, P. Földiák, Bayesian bin distribution inference and mutual information. IEEE Trans. Inf. Theory **51**(11), 3766–3779 (2005)
28. D. Endres, M. Oram, Feature extraction from spike trains with Bayesian binning: latency is where the signal starts. J. Comput. Neurosci. **29**, 149–169 (2010)
29. D. Endres, M. Oram, J. Schindelin, P. Földiák, Bayesian binning beats approximate alternatives: estimating peri-stimulus time histograms, in *Advances in Neural Information Processing Systems 20*, ed. by J. Platt, D. Koller, Y. Singer, S. Roweis (MIT Press, Cambridge, MA, 2008)
30. P. Fearnhead, Exact and efficient Bayesian inference for multiple changepoint problems. Stat. Comput. **16**(2), 203–213 (2006)
31. T. Flash, B. Hochner, Motor primitives in vertebrates and invertebrates. Curr. Opin. Neurobiol. **15**(6), 660–666 (2005)
32. T. Flash, N. Hogan, The coordination of arm movements: an experimentally confirmed mathematical model. J. Neurosci. **5**, 1688–1703 (1985)
33. E. Fuchs, T. Gruber, J. Nitschke, B. Sick, Online segmentation of time series based on polynomial least-squares approximations. IEEE Trans. Pattern Anal. Mach. Intell. **32**(12), 2232–2245 (2010)
34. A. Gensler, T. Gruber, B. Sick, Blazing fast time series segmentation based on update techniques for polynomial approximations, in *13th IEEE International Conference on Data Mining Workshops, ICDM Workshops, TX, USA, 7–10 December 2013*, pp. 1002–1011
35. M. Giese, A. Mukovskiy, A. Park, L. Omlor, J. Slotine, Real-time synthesis of body movements based on learned primitives, in *Statistical and Geometrical Approaches to Visual*

Motion Analysis, ed. by D. Cremers, B. Rosenhahn, A.L. Yuille (Springer, Heidelberg, 2009), pp. 107–127
36. R.D. Green, Spatial and temporal segmentation of continuous human motion from monocular video images, in *Proceedings of Image and Vision Computing*, pp. 163–169, New Zealand (2003)
37. C.M. Harris, D.M. Wolpert, Signal-dependent noise determines motor planning. Nature **394** (6695), 780–784 (1998)
38. C.B. Hart, S. Giszter, Distinguishing synchronous and time varying synergies using point process interval statistics: Motor primitives in frog and rat. Front. Comput. Neurosci. **7**, 52 (2013)
39. P.A. Højen-Sørensen, O. Winther, L.K. Hansen, Mean field approaches to independent component analysis. Neural Comput. **14**(4), 889–918 (2002)
40. M. Hutter, Exact Bayesian regression of piecewise constant functions. J. Bayesian Anal. **2**(4), 635–664 (2007)
41. W. Ilg, G. Bakir, J. Mezger, M. Giese, On the representation, learning and transfer of spatio-temporal movement characteristics. Int. J. Hum. Rob. **1**(4), 613–636 (2004)
42. Y.P. Ivanenko, G. Cappellini, N. Dominici, R.E. Poppele, F. Lacquaniti, Coordination of locomotion with voluntary movements in humans. J. Neurosci. **25**(31), 7238–7253 (2005)
43. Y.P. Ivanenko, R.E. Poppele, F. Lacquaniti, Five basic muscle activation patterns account for muscle activity during human locomotion. J. Physiol. **556**(Pt 1), 267–282 (2004)
44. T.R. Kaminski, The coupling between upper and lower extremity synergies during whole body reaching. Gait Posture **26**(2), 256–262 (2007)
45. E. Keogh, S. Chu, D. Hart, M. Pazzani, An online algorithm for segmenting time series, in *Proceedings IEEE International Conference on Data Mining, 2001*, *ICDM 2001,* pp. 289–296 (2001)
46. L. Kovar, M. Gleicher, F. Pighin, Motion graphs. ACM Trans. Graph. **21**, 473–482 (2002)
47. F. Kschischang, B. Frey, H.-A. Loeliger, Factor graphs and the sum-product algorithm. IEEE Trans. Inf. Theory **47**(2), 498–519 (2001)
48. D. Kulic, W. Takano, Y. Nakamura, Online segmentation and clustering from continuous observation of whole body motions. IEEE Trans. Rob. **25**(5), 1158–1166 (2009)
49. F. Lacquaniti, C. Terzuolo, P. Viviani, The law relating kinematic and figural aspects of drawing movements. Acta Psychol. **54**, 115–130 (1983)
50. D.D. Lee, H.S. Seung, Algorithms for non-negative matrix factorization, in *In NIPS*, pp. 556–562. MIT Press, Cambridge (2000)
51. D. Lemire, A better alternative to piecewise linear time series segmentation. CoRR arXiv:abs/cs/0605103 (2006)
52. Y. Li, T. Adal, V. Calhoun, Estimating the number of independent components for functional magnetic resonance imaging data. Hum. Brain Mapp. **28**(11), 1251–1266 (2007)
53. J. Liu, Y.C. Nakamura, N.S. Pollard, Annotating everyday grasps in action, in *Dance Notations and Robot Motion*, chapter ZZ, pp. XX–YY. (Springer, Berlin, 2015)
54. T. Minka, Automatic choice of dimensionality for PCA. Technical report, M.I.T. Media Laboratory Perceptual Computing Section (2000)
55. T. Minka, J. Winn, Gates: a graphical notation for mixture models, in *Proceedings of NIPS* (2008)
56. A. Mukovskiy, J.J.E. Slotine, M.A. Giese, Dynamically stable control of articulated crowds. J. Comput. Sci. **4**(4), 304–310 (2013)
57. L. Omlor, New methods for anechoic demixing with application to shift invariant feature extraction. Ph.D. in informatics, Universität Ulm. Fakultät für Ingenieurwissenschaften und Informatik, 2010. urn:nbn:de:bsz:289-vts-72431
58. L. Omlor, M. Giese, Blind source separation for over-determined delayed mixtures, in *Advances in Neural Information Processing Systems 19*, ed. by B. Schölkopf, J. Platt, T. Hoffman (MIT Press, Cambridge, MA, 2007), pp. 1049–1056
59. L. Omlor, M.A. Giese, Extraction of spatio-temporal primitives of emotional body expressions. Neurocomputing, **70**(10–12), 1938–1942 (2007) (reviewed)

60. L. Omlor, M.A. Giese, Anechoic blind source separation using wigner marginals. J. Mach. Learn. Res. **12**, 1111–1148 (2011)
61. F. Polyakov, E. Stark, R. Drori, M. Abeles, T. Flash, Parabolic movement primitives and cortical states: merging optimality with geometric invariance. Biol. Cybern. **100**(2), 159–184 (2009)
62. L. Rabiner, A tutorial on hidden Markov models and selected applications in speech recognition. Proc. IEEE **77**(2), 257–286 (1989)
63. C.L. Roether, L. Omlor, A. Christensen, M.A. Giese, Critical features for the perception of emotion from gait. J. Vision **9**(6), 1–32 (2009)
64. C. Rose, B. Bodenheimer, M.F. Cohen, Verbs and adverbs: Multidimensional motion interpolation using radial basis functions. IEEE Comput. Graph. Appl. **18**, 32–40 (1998)
65. A. Safonova, J.K. Hodgins, Construction and optimal search of interpolated motion graphs, in *ACM SIGGRAPH 2007 Papers*, SIGGRAPH'07, New York, NY, USA. (ACM, New York, 2007)
66. A. Safonova, J.K. Hodgins, N.S. Pollard, Synthesizing physically realistic human motion in low-dimensional, behavior-specific spaces. ACM Trans. Graph. **23**, 514–521 (2004)
67. M. Santello, M. Flanders, J.F. Soechting, Postural hand synergies for tool use. J. Neurosci. **18** (23), 10105–10115 (1998)
68. J. Segouat, A. Braffort, Toward the study of sign language coarticulation: methodology proposal, in *Proceedings of the 2009 Second International Conferences on Advances in Computer-Human Interactions*, ACHI '09, pp. 369–374, Washington, DC, USA, 2009. IEEE Computer Society
69. T.F. Shipley, M.J. Maguire, J. Brumberg, Segmentation of event paths. J. Vision, **4**(8) (2004)
70. A. Swindlehurst, Time delay and spatial signature estimation using known asynchronous signals. IEEE Trans. Signal Process. **46**, 449–462 (1997)
71. J.S. Thomas, D.M. Corcos, Z. Hasan, Kinematic and kinetic constraints on arm, trunk, and leg segments in target-reaching movements. J. Neurophysiol. **93**(1), 352–364 (2005)
72. L.H. Ting, J.M. Macpherson, A limited set of muscle synergies for force control during a postural task. J. Neurophysiol. **93**(1), 609–613 (2005)
73. G. Torres-Oviedo, J.M. Macpherson, L.H. Ting, Muscle synergy organization is robust across a variety of postural perturbations. J. Neurophysiol. **96**(3), 1530–1546 (2006)
74. S. Tu, L. Xu, An investigation of several typical model selection criteria for detecting the number of signals. Front. Electr. Electron. Eng. China **6**, 245–255 (2011)
75. A. Ude, C. Atkeson, M. Riley, Programming full-body movements for humanoid robots by observation. Robot. Auton. Syst. **47**, 93–108 (2004)
76. A. Ude, M. Riley, B. Nemec, T. Asfour, G. Cheng, Synthesizing goal-directed actions from a library of example movements, in *IEEE/RAS International Conference on Humanoid Robots (Humanoids)* (2007)
77. M. Wächter, S. Schulz, T. Asfour, E. Aksoy, F. Wörgötter, R. Dillmann, Action sequence reproduction based on automatic segmentation and object-action complexes, in *IEEE/RAS International Conference on Humanoid Robots (Humanoids)*, pp. 189–195 (2013)
78. L. Xu, Bayesian ying yang learning. Scholarpedia **2**(3), 1809 (2007)
79. Ö. Yilmaz, S. Rickard, Blind separation of speech mixtures via time-frequency masking. IEEE Trans. Signal Process. **52**, 1830–1847 (2004)
80. J.M. Zacks, T.S. Braver, M.A. Sheridan, D.I. Donaldson, A.Z. Snyder, J.M. Ollinger, R.L. Buckner, M.E. Raichle, Human brain activity time-locked to perceptual event boundaries. Nat. Neurosci. **4**(6), 651–655 (2001)
81. J.M. Zacks, S. Kumar, R.A. Abrams, R. Mehta, Using movement and intentions to understand human activity. Cognition **112**(2), 201–216 (2009)

Beyond Watching: Action Understanding by Humans and Implications for Motion Planning by Interacting Robots

Gowrishankar Ganesh and Tsuyoshi Ikegami

Abstract When you see an individual holding a baby in front of an elevator, you instinctively move to open the door for him. This seemingly obvious helping action is possible because you are able to immediately characterize, recognize and then *understand* his actions—that is, recognize the intention behind the observed individual's actions, estimate the outcome of his current actions, predict his future actions and infer his constraints due to the baby in his hands. Fast action recognition and action understanding abilities make humans adept at social interactions, and are fundamental requirements for future robots in order for them to interact with humans. While other chapters in this book focus on action characterization and recognition through the use of dance notations, in this chapter we will focus on the problem of understanding recognized actions. In particular, we aim to elucidate how ideas from action understanding research in primates can help robots formulate behavior plans when they interact with humans and other robots. We first briefly review the historical concepts, and psychological and neuro-scientific findings on action understanding by primates. Next, we detail the possible computational mechanisms underlying action understanding by humans. We then highlight the controversies regarding these beliefs and explain the results of our recent study that answers some of these controversies. Finally, utilizing results from our study, we propose and explain a conceptual bio-mimetic framework for action understanding by robots, in order to enable them to plan *helping* and *impeding* behaviors during interactions, similar to humans.

G. Ganesh (✉)
CNRS-AIST JRL (Joint Robotics Laboratory), UMI3218/CRT, Centre National de la Recherche Scientifique, Tsukuba Central 1, 1-1-1 Umezono, Tsukuba, Ibaraki 305-8560, Japan
e-mail: g.ganesh@aist.go.jp

T. Ikegami
Center for Information and Neural Networks, National Institute of Information and Communications Technology, 1-4 Yamadaoka, Suita City, Osaka 565-0871, Japan
e-mail: ikegmi244@gmail.com

J.-P. Laumond and N. Abe (eds.), *Dance Notations and Robot Motion*,
Springer Tracts in Advanced Robotics 111,
DOI 10.1007/978-3-319-25739-6_7

1 Introduction

Humans and robots can navigate their physical world using their senses, vision being predominant for the majority. However, when it comes to 'navigation' of one's social environment, sensory observations are usually not enough because as famously written by Plato, "things are not always what they seem". Taking an example from sports, consider that you are a soccer defender who sees an opposition player approaching with the ball. This visual information of the opponent and his surroundings alone is not enough for you to decide your next action. To decide what you should do next, you need to analyze the visual information and predict if the opponent intends to go left or right and whether he intends to pass the ball or try to dribble his way around you. In either case you need to predict the outcome (of the pass or dribble) in terms of the rough ball trajectory so that you can intercept it. The opponent's behavior will depend on his physiological characteristics like his size, speed and whether he is left or right footed, and on his mental characteristics, whether he is anxious, scared or has the confidence to pass you. The situation is complicated by the fact that the opponent is making similar predictions about your behavior, which means that you need to consider his predictions about you when you make your predictions about him!

The choice of behaviors in social interactions requires action characterization and recognition, for example by utilizing dance notations, followed by *action understanding*; deciphering the emotion and intention leading to an observed action, detecting constraints associated with the action and the performing agent, predicting the outcomes of the observed actions, and predicting future actions. In this chapter we aim to introduce the robotics community to action understanding research in primates, elucidate the computational beliefs, and show how these can be useful for the development of automatic interaction behaviors in robots. We are interested particularly in scenarios like our elevator and soccer examples where the interacting agents need to plan their behaviors individually and without any explicit information exchanges. We believe that such *individualistic* interaction planning scenarios define the bench mark for future *intelligent* robot-human interaction research.

But how to enable robots to understand humans, and other robots? An obvious strategy is to examine human interactions and then try to implement the same in robots, a procedure that has yielded promising automatic robot behaviors previously [1–3] and what we will utilize here. On the other hand, the mechanisms underlying action understanding abilities in humans are still not completely clear and a subject of ongoing research. Therefore with this chapter, while we present a conceptual framework that allows formulation of interaction plans by robots, we also want to motivate the requirement for continued integrated research in robotics and behavioral neuroscience in the future.

The chapter is organized as follows. Sections 2, 3 and 4 first provide a brief historical account of action understanding research starting from the philosophical motivations to the recent neuroscientific findings that clarify the computational mechanisms behind it. We conclude Sect. 4 by summarizing the controversy in regard to the previous findings. Section 5 discusses results from our recent study that answered some of these controversies. Finally, before going to the robotics discussions, in Sect. 6 we list the computational mechanisms in the neuroscience literature for action understanding. Section 7 summarizes previous action understanding research in robotics. Section 8 briefly defines some key concepts in *motion planning* and defines the planning problems during interaction. Section 9 then gives a description of our conceptual framework for planning robot behaviors during interactions with an example, before we conclude with Sect. 10.

2 Beyond Watching: The Philosophy Behind Action Understanding

Humans are adept at action understanding. Our social skills, from gestures for communication, sports, to driving a car safely, are critically dependent on our ability to understand observed actions performed by others and take appropriate actions ourselves [4]. Philosophers have questioned the mechanisms behind this human ability for a long time. The notion that actions are intrinsically linked to perception was first proposed by William James in his ideomotor principle. James claimed that "every mental representation of a movement awakens to some degree the actual movement which is its object" [5]. He also made similar observations between emotions and actions. He questioned the orthodox thinking that bodily expressions are the consequences of intrinsic emotions and proposed the converse, that in fact emotions were the consequences of bodily expressions. His proposal could, in principle, explain how a human can *feel* the emotions of a fellow individual, solely by observing his bodily expressions.

In order to explain aesthetic experience of an artwork by an observer, Theodor Lipps [6] borrowed the term *Einfühlung* (originally introduced by Robert Vischer and later translated in English as *empathy*) referring to the process by which the observer can imaginatively project himself into contemplated objects. Lipps later utilized *Einfühlung* to explain emotional perceptions of observed movements [6] suggesting that the feelings and intentions of a person can be perceived by projecting oneself inside him or 'walking in his shoes'. This concept was developed by Edmund Husserl who proposed that our ability to understand other agents stems from the belief that their bodily experiences are similar to our own experiences while acting on our bodies. The concept of empathy thus suggests how the perception of *self* and of others are coupled. Edith Stein extends this concept further in *On the Problem of Empathy* [7] (1912/1964, English translation), where she proposes the converse, that acts of empathy can help one to learn what type of person one is.

Similar closeness between perception of other's actions, and that of one's own were also expressed by the famous French philosopher Merleau-Ponty [8] when he said, "The sense of the gestures is not given, but understood, that is, recaptured by an act on the spectator's part. The whole difficulty is to conceive this act clearly without confusing it with a cognitive operation. The communication or comprehension of gestures come about through the reciprocity of my intentions and the gestures of others, of my gestures and intentions discernible in the conduct of other people. It is as if the other person's intention inhabited my body and mine his", and by Herbert Mead (1912) [9] who wrote, "Any gesture by which the individual can himself be affected as others are affected, and which therefore tends to call out in him a response as it would call out in another, will serve as a mechanism for the construction of a self".

3 From Philosophy to Psychology

The recurring theme in all the philosophical works is the relationship between the *other* and the *self*. The sense of *ownership* (the perception that an observed limb is part of your body), *agency* (the perception that a limb movement is produced by you) and presence (the perception that you are at a particular location), are probably the fundamental features defining the *self* and consequently, the *other* by dissociating from the self. Though the perception of self is critically determined by multi-sensory congruencies and perceived causality between the intended actions (or motor commands) and sensory perceptions, it can be very plastic and change quickly, as demonstrated by studies on tool-use [10, 11] and the now famous rubber hand illusion [12]. In this illusion, synchronous touches, applied to a rubber hand in full view of the participant, and the real hand hidden behind a screen, produce the sensation that the touches felt originate from the rubber hand, leading to a feeling of ownership of the artificial rubber hand [13]. For more details on the sense of ownership and agency readers can refer to [14–17].

The core idea behind the philosophy of action understanding is that the other is NOT very different from the self and, extending on Stein's beliefs, the two can in fact be interchangeable in the brain. This idea of closeness between self and other was extended into the domain of action-perception in the *common coding theory* introduced by Prinz [18]. This theory hypothesized a shared representation for both perception and action; seeing an event activates the action associated with that event, and performing an action activates the associated perceptual event. This hypothesis was later verified by the discovery of the *Mirror Neurons* (see next section).

4 Action Understanding by the Brain

4.1 Discovery of the Mirror Neurons

While the philosophical and psychological concepts suggested the role of one's own action production system (or motor system) in the perception of actions observed in others, a concrete proof for the existence of such a mechanism did not come until 1992 when Giacomo Rizzolatti and his colleagues discovered, what they called *Mirror Neurons,* in the brain of a Macaca nemestrina monkey [19] (Fig. 1). These neurons, isolated in a region associated generally with motor activations (i.e. in the motor system), exhibited the unique feature that they activate during both, the execution of an action and the observation of the same action performed by others. Mirror neurons were first discovered in the ventral premotor area (F5) [19] and then in the inferior parietal lobule (PF/PFG) [20, 21]. The cortical network consisting of these areas were named as the mirror system or mirror neuron system [22].

While the mirror system was initially discovered in monkeys, human brain imaging studies using Magnetoencephalography (MEG) [23–25], functional Magnetic Resonance Imaging (fMRI) [26–28], brain stimulation studies using

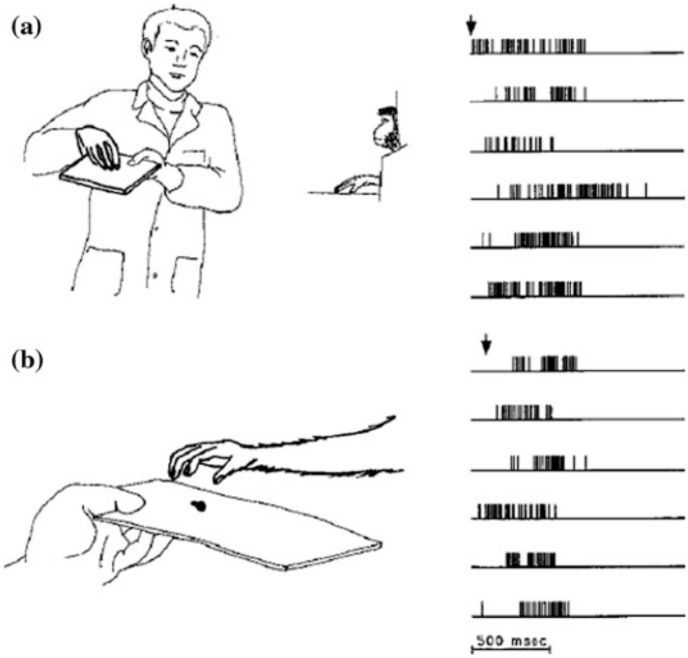


Fig. 1 In their seminal work [19], Giacomo Rizzolatti and colleagues discovered neurons in the ventral pre-motor cortex of a Macaca nemestrina monkey that activated both when, **a** the monkey observes a reach-grasp movement, and **b** when the monkey makes the same movement. Figure reprinted from [19] with permission

Transcranial Magnetic Stimulation (TMS) [29, 30], and a single unit recording study [31] have since provided sufficient evidence to show the presence of the mirror neuron system in the human brain. Interestingly, mirror neuron activity in the human brain has been observed to be modulated by the observer's own motor repertoire [32–35]. For example, the mirror system activities were observed to differ when viewing videos of classical ballet or capoeira depending on whether an observer was an expert in the observed task [33].

The discovery of the mirror neuron has exhibited a clear action specific coupling between the visual and the motor system that has motivated a series of debates, and controversies [36–39] regarding the functional significance of these neurons in regard to various social abilities in humans, including verbal communication, Theory of Mind and action understanding. We will concentrate on the investigations regarding the role of mirror neurons in action understanding. Readers can refer to [22, 40, 41, 42] for a discussion on the role of mirror neurons in other social functions.

4.2 Mirror Neurons, Motor System and Action Understanding

Does the brain utilize the motor system, namely the mirror neurons, to understand observed actions? Multiple reports from monkey and human studies suggest this to be true.

In their electrophysiological study, Umilta et al. [43] first isolated mirror neurons that activated both during the execution of a reach-grasp motion and the observation of the reach-grasp made by an experimenter. Next they exhibited that part of these mirror neurons activated only when the object, which was the goal of the task, was present and not when the same reach-grasp movement was mimed. Critically the neurons activated even when the grasped object was not visible (blocked with a screen) to the monkey but the monkey knew the object was present behind the screen. The authors claimed that this observation exhibits that the mirror neurons code not just the kinematics of the observed action, but the goal of the observed action. Other monkey electrophysiological studies have similarly exhibited motor activations related to action intention [20] and sounds [44] associated with actions.

In humans, fMRI [45, 46] and TMS studies [47–49] have shown that the mirror or motor system responding to the same observed action in terms of kinematics, are modulated by the contexts or prior information indicating what (goal) and why (intention) the agent is doing the action. For example, Iacoboni et al. had subjects observe pictures of an agent grasping a cup in different contexts reflecting different intention of the action. The "before tea" context suggested grasping the cup for drinking, while the "after tea" context suggested it is for washing the cup. They observed that frontal motor areas, that were previously associated with mirror neurons, exhibited significantly higher responses to the actions performed with context compared to actions without context, or to the presentation of static scenes suggesting the contexts [45].

Furthermore, behavioral studies have shown that the human ability to understand actions performed by others correlates with their ability to execute the same actions. In their study with developing infants, Kanakogi and Itakura observed a correlation between an infant's ability to make grasping movements and their ability to make predictive eye movements while observing reaching movements performed by others [50]. Other studies have reported the ability to predict action goals by infants [51] and action outcomes by professional sportman [35] to depend on their motor abilities.

4.3 Action Understanding by Association

Before moving on to the criticisms on the role of the motor system in action understanding, we note here that although in this chapter we motivate and concentrate on the role of the motor system in action understanding, this is not the only mechanism by which action understanding is enabled. There exists another, widely accepted mechanism of action understanding, *Associative learning* (also referred to as "understanding from outside" [42] or "understanding-by-association" [52]).

Associative learning refers to the process where an individual can understand actions, for example predict the outcome of an action, simply by learning (and remembering) the outcomes from when he previously observed the same actions [53]. Associative learning can explain how patients with lesions or damages in the motor areas and who are unable to make a repertoire of actions still can understand the same actions performed by others [36, 37] and why we can understand and predict the actions by non-human agents like pet animals or machines [54], even though obviously our motor system is very different from theirs. On the other hand, there is evidence to support that motor abilities in a task can additionally improve understanding developed through associative learning. In their experiment with professional basketball players, Agliotti and colleagues observed that athletes predicted the success of free shots at a basket earlier and more accurately than coaches or sports journalists, who can be considered to have comparable visual experience and hence comparative associative learning [35]. Therefore, arguably our ability to understand actions stems from the interactions of both the associative and motor mechanisms.

4.4 Criticism Against Motor Role in Action Understanding

However, in spite of a multitude of evidence supporting the role of the motor system in action understanding, this issue is still not popularly accepted and remains highly controversial [36, 37, 38, 39, 53, 55]. The controversy arises because the conclusions from the monkey electrophysiology [20, 43], human brain imaging [45] and child development studies [50, 51] in this regard (that we have presented in the last sections) are considered correlative and not causative in nature [37–39]. The issue is further complicated by the fact that it is very difficult to quantify action understanding in monkeys. On the other hand, the results from lesion [56] and brain stimulation [29, 48, 49] studies have been criticized to be inconclusive because it is difficult to concretely access the functional extent of a lesion or electrical stimulation, which may include both motor and action understanding neurons [22]. Indeed no previous study has exhibited a causal relation between action understanding and action production, where a purposely induced change in the action understanding system affects action production, or vice versa.

In our recent study [57] we were able to present the first direct causal relationship between action production (by the motor system) and outcome prediction, a component of action understanding. We utilized a novel behavioral *outcome prediction learning* paradigm for this purpose, where the understanding of an individual is changed not through neural intervention but through learning. This paradigm enabled us to induce a focused change in the understanding system of individuals while avoiding the spatio-functional ambiguities (and the previous controversy) associated with neural and lesion studies, quantify the understanding change and then observe its effects on the individual's motor performance. We detail the results of this study in the next section.

5 Causation Between Action Production and Outcome Prediction of Observed Actions by Humans

5.1 Experiment Paradigm

Our study included 21 darts 'experts', three first time dart throwers (novice), and three novice ten-pin bowlers across three experiments. The experiments utilized the outcome prediction learning paradigm and examined the behavioral interaction between action production and outcome prediction [35, 50, 58], which is considered as a component of action understanding [59, 60]. Specifically, we asked expert dart throwers to predict the outcome of throws made by an unfamiliar darts novice by watching the novice darts player's throwing action. We regulated the relevant *prediction error* feedbacks available to the experts, controlled the improvement in their prediction ability [60, 61] and exhibit that this affects the accuracy of the expert's own dart throws.

Behavioral paradigms examining interference and transfer of learning between tasks have been previously utilized to investigate the neural processes behind human motor learning [62–68]. Here we use a similar procedure for *outcome prediction learning*. Behavioral paradigms cannot measure or identify where neural activity related to a behavioral function takes place in the brain. However their advantage lies in the fact that with proper control, they can ensure changes in neural processes specific to a behavioral function wherever they lie in the brain. For our purpose, the outcome prediction learning paradigm enabled us to induce targeted changes in the outcome prediction system of individuals while avoiding the spatio-functional ambiguities characteristic of changes induced by lesions [56] and neural interventions [29, 48, 49]. We chose to use darts experts as subjects due to several reasons: (i) Experts in a sport are known to possess an excellent ability to predict the outcome of observed actions [35], (ii) Arguably, the observing expert will not explicitly imitate the novice and, (iii) an expert's motor performance is expected to be stable with time and resistant to fatigue. We could thus exclude any major contribution of explicit strategy changes [69] and fatigue in our results.

5.2 *Experiment-1: Watching Darts Versus Watching Bowling*

Experiment-1 extended over two days. 16 darts experts threw 70 darts (aimed for the center of the darts board) each day over two visual feedback (VF) blocks, where they could see where their darts landed on the board, and five blocks without visual feedback (nVF) where the room light was switched off when they threw their dart so they could not see where their darts landed (Fig. 2a). The nVF blocks were

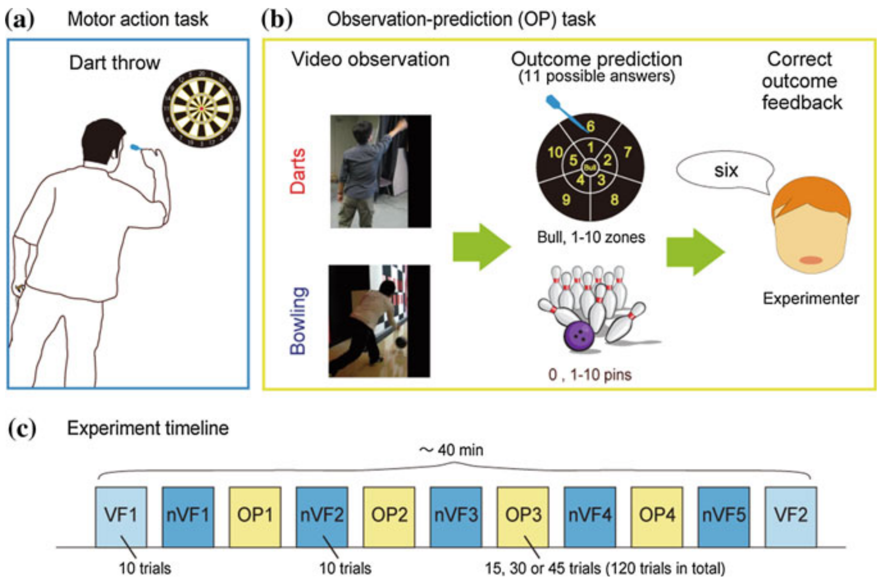


Fig. 2 Our experiment consisted of **a** two motor action tasks, one in which the subjects threw darts in the presence of visual feedback (*VF*) of where their darts land on the darts board and second, in the absence of visual feedback (*nVF*), and **b** observation-prediction (*OP*) tasks in which the subjects observed the video of either a novice darts thrower or a ten-pin bowler (snap shots shown), made a prediction of the outcome of each throw, and were given the feedback of the correct outcome orally by the experimenter. The chance level for both OP tasks was 9.09 % (=1/11 × 100). Each experiment session followed the sequence of blocks as shown in **c.**

interleaved with observation-prediction (OP) blocks, where the experts watched the video of a novice darts thrower (on one day) or a novice ten-pin bowler (another day) as control. Part of the videos in both cases were masked such that the dart flight trajectory and darts board, and the bowling ball trajectory and bowling pins were not visible to the viewers (Fig. 2b). Novice subjects were asked not to show any expressions after their throws and the recorded video was further checked and edited to remove any throws that still contained some expressions after. The experts were informed of the 'goal' of the novice actions (hitting the board center or 'bull' in case of dart throwers and felling all ten pins or a 'strike' for bowlers) and asked to predict the outcome of the throws (in terms of either the location on a lower resolution darts board or the number of bowling pins felled) by watching the action kinematics in the videos. They were informed of the correct outcome orally after each prediction. The experiment time line is presented in Fig. 2c. The nVF blocks were used to prevent visual correction by the experts and helped in magnifying the effects of the OP task on their darts performance.

The outcome prediction in the experts improved significantly through the OP task of Experiment-1 both, when they watched the video of a novice ten-pin bowler (abscissa of red plot in Fig. 3b; 8.57 ± 4.15 SD % correct predictions above chance, t(15) = 8.25, $p < 0.001$) and a dart thrower (abscissa of blue plot in Fig. 3b; 11.69 ± 7.15 SD, t(15) = 6.54, $p < 0.001$). Correspondingly, a two-way ANOVA revealed significant interaction ($F_{1,15} = 9.23$, $p < 0.01$) between the darts performance across VF block (VF1, VF2) and the observed video in the OP task (darts prediction, bowling prediction). Though the initial performance of experts was similar in the VF1 block ($F_{1,30} = 3.07$, $p > 0.05$), a significant increase in the performance error was observed in experts when they watched a darts novice ($F_{1,30} = 15.55$, $p < 0.001$) but not when they viewed a bowling novice. The darts performance deterioration, defined as the increase of performance error between VF1 and VF2, was therefore significantly positive in Experiment-1 (Ordinate of red plot in Fig. 3b; t(15) = 5.10, $p < 0.001$).

Experiment-1 thus exhibited two results. First, predicting a novice's action leads to a progressive increase in the performance error in the expert dart throwers. Second, the performance change is task specific: darts performance error increases on predicting outcomes of darts throws but not on predicting outcomes of ten-pin bowling, critically even when the improvement in outcome prediction was similar between darts and bowling OP task conditions (t(15) = 1.22, $p = 0.24$). The absence of performance changes in the bowling sessions (*blue plots*, Fig. 3b) shows that the increase in performance error is not due to fatigue, loss of attention or motivation or lack of visual feedback.

However, while Experiment-1 exhibits that watching novice dart throwers deteriorates the performance of experts, it does not conclusively exhibit that the deterioration is due to changes in the outcome prediction system. The outcome prediction did significantly improve in Experiment-1 but, the performance deterioration may have been unrelated to this prediction change and may have resulted simply due to unconscious mimicry (related to the so called Chameleon effect [70] of the observed novice's darts action which was different, both in style and

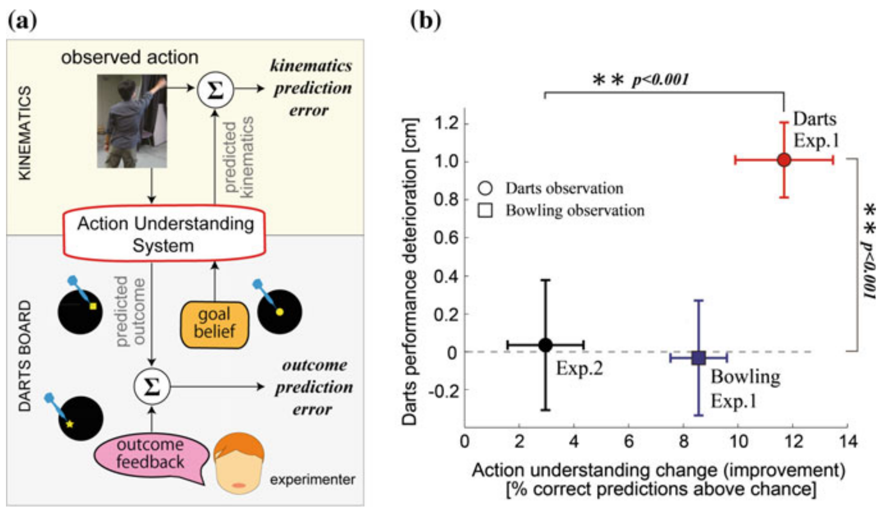


Fig. 3 **a** The experts could utilize two types of prediction errors to improve their outcome prediction in our task. First, the outcome prediction error between the expert predicted outcome and the correct *outcome feedback* from the experimenter. Second, the kinematic prediction error between the action kinematics predicted by the expert corresponding to his *goal belief* (of where the novice aims his throws for), and the kinematics the expert actually observes in the video. We modulated the outcome feedbacks and goal belief provided to the expert subjects across our experiments. **b** In the presence of both the prediction errors in Experiment-1 (*red plot*), the outcome prediction change (*abscissa* of *red plot*) was significant, leading to darts performance deterioration (*ordinate* of *red plot*). On the other hand when the experts watched the bowling videos, the outcome prediction change did not affect the darts performance (*blue plot*). When the outcome feedbacks and goal belief were both removed in Experiment-2, the outcome prediction change (abscissa of *black plot*) as well as the performance deterioration (ordinate of *black plot*) were prevented

variability, in comparison to the expert's. To exhibit that the improvement in the outcome prediction is indeed the cause of the performance deterioration, we conducted an additional experiment (Experiment-2) and examined how the performance deterioration is affected when the improvement in outcome prediction of the observed darts action is modulated by us.

5.3 Prediction Errors for Outcome Prediction Learning

The experts could utilize two *prediction errors* to improve their outcome prediction [60, 61, 71, 72] in the OP blocks of Experiment-1. The first is the outcome prediction error—the difference between the outcome predicted by the expert from the observed novice action, and the correct outcome provided to him orally by the experimenter (Fig. 3a). Second is the kinematics prediction error—the difference between the kinematics expected by the expert corresponding to the goal he believed the novice aimed for (the center of the board), and the novice kinematics he actually observed (Fig. 3a).

5.4 Experiment-2: Watching Darts Without Prediction Errors

In Experiment-2, 16 experts (11 from Experiment-1 and 5 new recruits) were again asked to watch and predict dart videos (a different novice's video was used for experts who participated in Experiment-1) in the OP task. We removed the two types of prediction errors, expecting this to completely suppress the improvement of outcome prediction in the darts experts. The outcome prediction error was removed by removing the feedback of the correct outcome provided to the experts. On the other hand, the kinematics prediction error was suppressed by removing the expert's goal belief. We mis-informed the expert at the start of the experiment that "the novice does not always aim for the center but aims for unknown targets provided by us and that we display only those trials in which he was successful". We expected the mis-information to remove any prior goal belief that the expert may have. As expected, in the absence of prediction errors, the outcome prediction in Experiment-2 (black plot in Fig. 3b) was significantly lower than in Experiment-1 ($t(29) = 3.82$, $p < 0.001$) and not different from chance. The outcome prediction system was thus little affected in Experiment-2. Importantly, in contrast to Experiment-1, there was no evidence of performance deterioration in Experiment-2 (Fig. 3b, black plot; $t(15) = 0.11$, $p = 0.92$). Note that except for the removal of the prediction errors in the OP task, all other conditions, including the observed darts novice videos and the initial level of darts performance (evaluated as the darts error in VF1; $t(29) = 1.91$, $p = 0.19$), were same between Experiment-1 and Experiment-2. Therefore, clearly the improvement in outcome prediction was the cause of the performance deterioration in Experiment-1.

To summarize, across our experiments we observed that a focused change in the outcome prediction ability led to a modification in the motor performance in expert darts player. While, these behavioral results prevents us from making conclusions about the role of the mirror neuron system, they clearly demonstrate a causal relation between the action production and outcome prediction of observed actions by humans, and provide strong support that at least part of the motor system is involved in action understanding by humans.

6 Computational Mechanisms of Action Understanding

Up till now in this chapter, we have discussed how the concepts in action understanding, and specifically the role of the motor system in action understanding, were developed in philosophy, extended into the domain of action-perception in psychology and are being supported by neuroscientific evidence from brain imaging, brain stimulation and behavioral studies. Before we can go on to discuss the implications of these concepts and results for robotics, we need to answer one key issue that will enable us to bridge the gap between neuroscience and robotics—How can the motor system, or what computations by the motor system, enable understanding of observed actions?

The prominent ideas in this regard can be categorized into three categories: Direct Matching mechanism (DMM), Hierarchical Predictive Coding (HPC) and Movement simulation (MS).

6.1 Direct Matching Mechanism (DMM)

The direct matching mechanism is reminiscent of the common coding theory and was proposed by Rizzolatti and his colleagues [22, 40]. In this framework, observing actions performed by others automatically triggers "resonated" activation of the observers' motor system which is required for the execution of the same action thus enabling imitation naturally. Furthermore, DMM proposes that due to the similarity between the action and perception activations, an observer can decode the intention, goals and outcomes associated with an action "from inside" (that is, based on one's own motor process). The closest robotics parallel of DMM is probably the application of affordance theory [73] in robotics [74].

6.2 Hierarchical Predictive Coding (HPC)

Other researchers have argued that DMM is not versatile enough to explain the understanding of many daily actions which are kinematically similar but lead to different outcomes. For example, a swinging arm can represent a person hailing a taxi or, swatting a wasp [60, 75]. In order to deal with such cases Friston and colleagues proposed a hierarchical action understanding schema which they referred to as the Hierarchical predictive coding framework (HPC) [76, 77]. In this schema, action understanding is achieved through hierarchically organized processes, each *understanding* the action at a different level- at the level of intention, goal, kinematics, and outcomes for example. The motor transformations between action commands and their consequences form the lowest level of this framework. Information processed at each upper level is transmitted as a prior for the predictions at a lower level while the prediction errors at each lower level are used for bottom-up update of information in the immediately upper level. An input to any level of the system thus leads to a cascade of self-organization until the prediction errors converge to zero, and consequently the observed action is *understood*. In this framework, taking the example of the swinging arm, higher level processing of the behavioral context (e.g. an agent is doing the action in his garden) can enables the observer to infer the most probable intention behind the observed action (that is, swatting a wasp).

6.3 Movement Simulation

The *movement simulation* architecture is motivated from studies in human motor control where it is generally agreed that human movements are enabled by pairs of *internal forward* and *inverse* models [78–80] of the environment and self. These models are learnt through experience and store different features of the dynamics defining an environment or one's own physical system. The forward models predict the sensory consequence of motor outputs and are utilized for perception of self-action [81–84], for online motor control [85–87], motor learning [88, 89] and in the identification of behavioral context [90, 91]. On the other hand the inverse model is believed to develop motor commands given certain sensory state feedbacks [78, 79, 92].

In an inter-personal context, this theory believes that the observed actions can be understood by simulating them in one's own brain using the same (or similar) internal models [93–96]. Kinematic information of the observed action may be sent to an observers' inverse model which transforms it into a motor command which the observer would have generated to achieve the same observed kinematics. The generated motor command can then be sent to the observer's forward model to enable the observer to predict the consequence of the observed action. If the observer is considered to have multiple forward inverse model pairs, one for each of the many possible contexts like in the MOSAIC architecture [90, 96], then each module can make its own prediction of the consequence. The observer can make a final prediction by weighting the prediction from each module by his prior knowledge of the context, while a comparison of the individual module predictions with the actual consequence can enable him to update the understanding of the context to best explains the observed action (see Wolpert et al. [96]).

Though conceptually similar, note that this simulation is different from the simulation theory of empathy [97, 98] in the so called field of Theory of Mind which deals with mental states. Though, a hierarchical implementation of the MOSAIC model that is said to allow understanding of the intention and goal behind observed action [96, 99] brings the two concepts in motor control and social neuroscience close.

7 Action Understanding in Robotics: A Short Summary

The concept of action understanding is not new in robotics and has been previously explored in detail, particularly in the field of imitation learning by humanoid robots [100, 101]. Depending on the task, a robot may be required to imitate a motion, action goal or action intention. Even when the aim of a robot is only to imitate the motion of a demonstrator (for example, a robot imitating a dance motion), a robot requires to first account for the viewpoint transformation of the observed motion, use the appropriate sensory-motor matching [101–103] and then solve the *body correspondence problem* [104, 105], to account for the differences in kinematics (and even dynamics) between the robot and the demonstrator. The requirement of action understanding is more intuitive when the aim of a robot is to imitate the goal or intention (e.g. [106, 107]). For example, imitating a hand reaching for a cup is a goal directed imitation that requires the robot to understand that the goal of the task is not to move the hand but pick up the cup. Imitation of a demonstrator's hand movements while he holds a big box with both his hand may be an example of an intention directed imitation where the robot needs to understand that the robot has to not just move the hands like the demonstrator but also keep a hold on the box while doing so.

Various mathematical strategies have been used for estimating the goal and/or intention for imitation. These can be summarized as methodologies that rely on a pre-learnt map (e.g. [108, 109]) that may be considered as an implementation of the direct matching mechanism, probabilistic models like in [110, 111] that may be considered to be close to hierarchical predictive coding, and probably the most popular, models utilizing Inverse optimization, inverse optimal control, inverse planning and inverse reinforcement learning [112–115] that can be considered to be movement simulation.

8 Motion Planning for Interacting Robots

Interestingly, except imitation learning, as far as we know, there is little implementation of action understanding in any other forms of interaction behavior by robots. Through the next two sections, we will present our conceptual idea with examples to show how action understanding, and specifically movement simulation, can play a key role in the planning of behavior in robots that intend to *help* or *impede* humans, and other robots.

8.1 Motion Planning in Robotics

Motion planning in a well-established and widely researched field, and interested readers are requested to refer to [116, 117] for an in depth discussion on it. Here we will not discuss motion planning in detail but only define some key concepts and definitions that will aid us in explaining our concepts and examples in the following sections.

In classical mechanics, the vector space required to define the states (or configuration) of a system are called its *configuration space,* while the set of actual configurations possible under real life constraints is called the system's

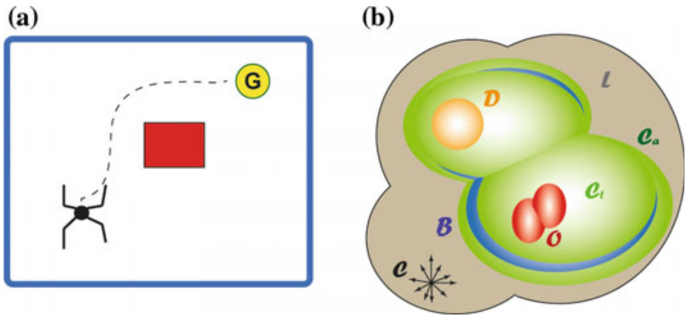


Fig. 4 **a** An example of an insect robot in constrained environment, and **b** the cartoon of the corresponding constraint representations in the configuration space

c*onfiguration manifold.* Taking a concrete example, consider the two dimensional, quadrapedal robot insect in Fig. 4 that can move inside a given blue boundary. Suppose the robot has a spherical point body connected to four legs each with a hip and knee joint. Then the dimension of the robot's configuration space $\boldsymbol{\mathcal{C}}$ will be $\mathbb{R}^2 \times \mathbb{S}^8 = 10$ (2 translational for the body and 8 rotational for the legs) with S representing a unit circle. Now considering that the robot movement is restricted by the boundary $\boldsymbol{\mathcal{B}} = \mathbb{R}^2$ and suppose the knee joints are constrained by the joint limits $\boldsymbol{\ell} = \mathbb{S}^4$, the actual configuration space that the robot can traverse can be given by $\boldsymbol{\mathcal{C}_a} = \mathbb{R}^2 \times \mathbb{S}^8$ (in Fig. 4b). Furthermore, within this actual configuration space there may be constraints that limit the robot behavior. We can have *volume reducing constraints*, like obstacles $\boldsymbol{\mathcal{O}}$ which can usually be defined by inequalities $\mathrm{O}(\boldsymbol{c}) < 0, \boldsymbol{c} \in \boldsymbol{\mathcal{C}_a}$. We can have *dimension reducing constraints* $\boldsymbol{\mathcal{D}}$ like when the robot cannot use all its legs near the wall boundaries, and which can be represented by equalities of the form $\mathrm{D}(\boldsymbol{c}) = 0, \boldsymbol{c} \in \boldsymbol{\mathcal{C}_a}$. Then if the task of the robot is to move to goal *G*, the remaining task configuration manifold $\boldsymbol{\mathcal{C}_t}$ (Fig. 4b) defines the configurations that the robot can traverse while moving to goal *G*. Here we defined the task configuration space for a relatively simple case but a similar configuration space can be defined for any complex system and task. Given this mathematic space, the goal of a motion planner is usually to find the best path to particular goal configurations (or their functions) while minimizing a predefined criteria.

Popularly, two categories of methods have been employed to solve planning problems in robotics, and especially humanoid robotics [118]. First *Prioritized Inverse Kinematic* strategies, which try to take advantage of the redundancy in the robots to assign multiple tasks to completely or partially decoupled subsets of the configuration space of the system in a prioritized manner (e.g. [119–123]). And second, *Configuration Space Exploring* strategies that explore the task configuration manifold to find the *optimal solution* that minimizes a given criterion or *edge cost* [117]. In the case of high dimensional systems like a humanoid robot, these explorations may be sparsified utilizing algorithms such as *Probabilistic Roadmaps* [124] or *Rapidly exploring Random Trees* (RRT) [125]. See [117] for an extensive review.

8.2 Challenges to Planning Inter-Agent Interactions

Inter agent interaction and collaborative task planning has been extensively researched in the field of swarm robotics [126, 127]. However, here we are particularly interested in behavior planning of robots that will interact with humans. For example, in our soccer scenario from the beginning of the chapter, our interest would be to plan the behavior for a robotic soccer defender who requires to tackle a human opponent. This problem differs from most of the swarm robotics scenarios in two critical aspects:

1. Most swarm robotics algorithms utilize centralized or hybrid (a combination of centralized and de-centralized) control where there is a 'big boss' computer that can control and observe every individual interacting agent. On the other hand when interacting with a human, it is not possible to have a centralized controller or observer to help the robot. Robots that interact with humans require individual, de-centralized planners.
2. Even the algorithms that propose de-centralized control in swarm robotics, utilize communication channels where agents can explicitly communicate their *utility functions* (for example their capacity to do a particular task) to one another so that they can collectively make decisions. In case of human-robot interactions, explicit communication between agents is improbable in most scenarios, and the robot would require to *learn* and utilize subtle physical and haptic cues [128] to plan its actions.

So to repeat from the beginning of the chapter, in order to be a successful soccer defender our robot would require to decipher the intention (and maybe emotion) behind the human opponent's action, detect constraints associated with the human's action, predict the outcomes of his current actions, and predict future actions, and decipher all these issues without any communication with the human. In other words the robot would require action understanding.

9 Our Proposed Bio-mimetic Solution

9.1 An Example Interaction Task

In this section we propose a framework to enable action understanding by robots. We will utilize features of human action understanding, and particularly movement simulation to formulate and develop *helping* and *impeding* behaviors by robots during interactions with another agent. For sake of clarity, we will take a simple 2-dimensional, two robot example to explain our concept and discuss the extension of this concept to interaction with humans in the next section.

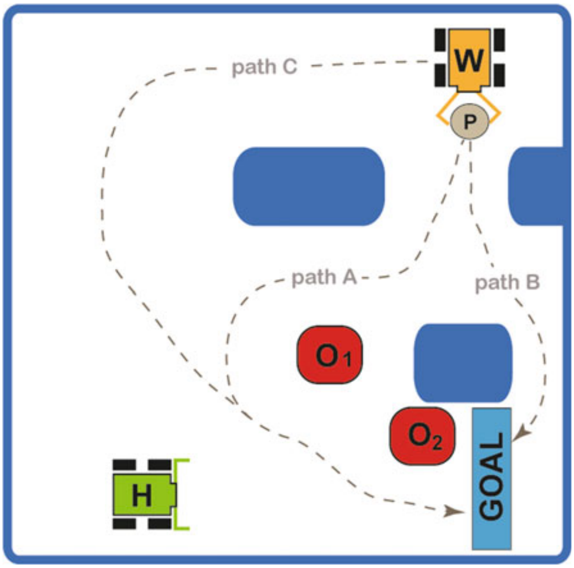


Fig. 5 Our 2-dimensional example involves two mobile robots that move inside a region bonded by blue non-movable walls. Robot *W* aims to bring peg *P* to the *GOAL* while robot *H* aims to *help* robot *W* in its task. O_1 and O_2 are manipulable obstacles that *W* is unable to move because it is holding the peg. The question we answer is how can *H* decide an action that would help *W*?

Let us consider a scenario with two interacting mobile robots, worker robot *W* and a helper robot *H* (Fig. 5), both with pincers with which they can hold and manipulate objects. We will assume that the robots can rotate arbitrarily and immediately about their base such that the robot's configuration space can be considered as $\mathcal{C}_a = \mathbb{R}^2$ (and in fact will look very much like Fig. 5) and that they move at a constant speed. We also assume that the robots are aware that the red objects (O_1 and O_2) are manipulable while the blue walls are not. Finally, to disregard the effect of time on robot *H*'s behavior, we assume that robot *H* can move much faster than robot *W*.

The task of robot *W* is to get the peg *P* to the *GOAL* in the shortest possible time and the task of robot *H* is to *help* robot *W*. Our goal here is to identify and execute *help*. In order to help robot *W*, robot *H* has to achieve the following objectives:

1. It has to identify the task aim of robot *W*; that it wants to take *P* to *GOAL* as fast as possible.
2. It has to estimate the physical constraints of robot *W* that arise from its size and shape.
3. It has to estimate the task constraints of robot *W*; that *W* has to hold the peg and hence cannot manipulate other objects in the environment without wasting time.
4. It has to estimate what path robot *W* will take.
5. Considering 1–4, decide what manipulation in the environment can help robot *W*.
6. And finally, plan an action to execute the helping operation.

9.2 Concept and Algorithm

We will start by omitting (1) here because in most applications the general aim of an agent remains fairly constant (for example the aim of players throughout a soccer game is to score as many goals as possible). We assume the aim of robot *W* is known to Robot *H*. As mentioned before, our bio-mimetic solution to the problem is motivated by motor simulation. We propose that *at the start of the task robot H should assume that robot W is similar to itself and consider itself to be doing robot W's task*. By making this simple assumption, *H* can solve objectives (2), (3) and (4) by movement simulation. That is, robot *H* can assume robot *W's* physical constraints to be same as its own. It can define the constraints due to the peg by assuming the peg is being held by it (robot *H*). Robot *H* can then combine these constraint estimations with the environment information (volume constraints because of the walls and objects O_1, O_2 in our case) and perform *simulation planning;* develop a motion plan that *it* (robot *H*) would take to complete the task, and then assume that robot *W* would do the same. Robot *H* can thus estimate that robot *W* will take say *path A* (in Fig. 5) associated with a cost V_a (time or path length for our case). V_a at any path point gives the cost of moving from this point to the *GOAL* along *path A*.

Once robot *W*'s path is identified, we are left with the last question of identifying and planning the help. The generalized solution for the help identification and planning would be to simulate plans multiple times by *moving* each manipulable constraint (objects O_1 and O_2 in our case) within their configuration manifold in the environment and calculating the change in the cost in each case in order to identify the manipulation that would constitute the greatest cost reduction, and hence the best help.

The generalized solution can of course be computationally intensive. However, simpler algorithms can be defined for specific help behaviors. For example we propose a configuration space–distance based algorithm that can be useful for help planning in many common tasks. The algorithm involves two steps: help identification and help manipulation.

Help identification

Step 1 We start by first down sampling the configuration space path that we expect W to take (*path A*), and the associated cost, into n *path points*.

$$\boldsymbol{p}_{\mathrm{i}}, \boldsymbol{p} \in \boldsymbol{\mathcal{C}}_{\boldsymbol{a}}. \wedge \mathrm{i} = [1, \mathrm{n}]$$
$$\boldsymbol{v}_{\mathrm{i}}, \boldsymbol{v} \in \boldsymbol{V}_{.} \wedge \mathrm{i} = [1, \mathrm{n}]$$

Step 2 Next we neglect all manipulable volume and (if any) dimensional constraints in our task and define the *unconstrained configuration space*. Then between each pair of path points, we find a difference of cost when moving from one point to the other traversing the planned path, and when traversing the shortest path in the unconstrained configuration space.

Set counter→0
For i=[1,n]
For k=[i+1,n]
counter→counter+1
Costdifference(counter) $= (\boldsymbol{v}_i - \boldsymbol{v}_{k)} - (\boldsymbol{v}_i - \boldsymbol{v}'_{ik})$
Loop k
Loop i

$\boldsymbol{v}'_{\mathrm{ik}}$ represents the cost at $\boldsymbol{p}_{\mathrm{k}}$ when $\boldsymbol{p}_{\mathrm{k}}$ is reached from $\boldsymbol{p}_{\mathrm{i}}$ in the unconstrained configuration space.

Step 3 We sort the cost differences in the decreasing order and starting from the top, consider the constraint associated with each difference and identify one (or more if allowed) as the *help constraint* to be manipulated during the helping action.

Help manipulation

If obstacle removal is the role the helper aims for, then the helper can move the identified obstacle utilizing the cost gradients near the object location in the (unconstrained) configuration space. The obstacle needs to be moved to a location where the cost is high because the robot will avoid this location.

Note that the same cost difference algorithm can be modified to not 'identify a constraint' but 'identify a point in the configuration space to be warped'. That is, once the help constraint is identified, the manipulation is not to move the constraint obstacle but move the two path points (between which the constraint exists) towards one another. If the *GOAL* is movable, this procedure will make the robot *H* help robot *W* by bringing the *GOAL* towards *robot W*.

9.3 Correcting for Estimation Errors

While the movement simulation by robot *H* serves as an initial estimate of robot *W*'s behavior, *H* can improve and correct these estimates by observing the actual behaviors by *W*. For example, if suppose robot *H* is half its present size in Fig. 5 then, while making the initial assumptions about the physical constraints, *H* will wrongly assume *W* to be as small as itself and expect *W* to take *path B* instead of *path A*. However, once robot *W* starts to move, *H* can notice that *W* (which follows *path A*) is not following *path B* and thus make corrections to its estimate of *robot W's* constraints. While in this simple example, this adjustment is trivial, in more complicated cases the mapping between these *estimation errors* and the planning constraints can be one to many. Depending on the nature of the agents, some situations may allow for simple scaling solutions while others may require priors or comparison of multiple models similar to the MOSAIC [90] architecture to enable corrections.

The concepts explained above have been represented graphically in Fig. 6.

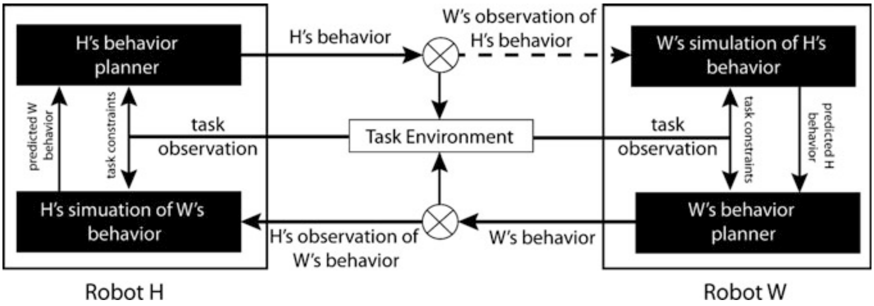


Fig. 6 A graphical representation of our proposed conceptual robot interaction planning framework (motivated from [129]). The dashed trace «W's observation of H's behavior» introduces inference looping between the agents

9.4 Challenges for Human-Robot Interactions

A big challenge for human interactions is the inference of the human intention during decision tasks. Motor simulation can help intention inference by letting a robot make an initial intention prediction from its own motor repotoire and then modify it (if it is wrong) with either the observed estimation errors or by observing responsive actions, which are relatively easier to understand. This procedure is similar to what humans do. For example, consider that you are sitting on a crowded train and see an old man approaching. You may first assume that he intends to sit and thus start to get up from your seat. However, you will change this assumption if you see him walking towards the door, or if he gestures you to keep sitting.

However human gestures and social behaviors can change extensively with different cultures and across individuals. Therefore in addition to action understanding by movement simulation, robots would also need to incorporate an associative learning-based action understanding (Sect. 4.3) procedure where behaviors experienced with a person (or in a particular culture) are stored in memory to be utilized for subsequent interactions, especially in regard to estimation error corrections. The memory requirement for this purpose can be large and movement notations, that can help store whole movements with minimal memory, will play an essential role in this aspect.

Once the intention is estimated, at least in the case of human-humanoid robot interactions where the physical dimensions of most humanoids are comparable or easily scalable to that of humans, simulation can quickly provide a robot with a good estimate of a human's configuration space and the volume reducing and dimension reducing constraints in a task. The robot can then predict the human behavior plan in most tasks by utilizing well established *human* cost functions [67, 130, 131], and thus plan how to help him. Online modifications (utilizing estimation error) may again be required in some cases at this stage to correct for errors in the assumed cost or due to the choice of sub optimal samples (as random sampling is utilized for planning in high dimensional spaces, see Sect. 8.1) in the configuration space during the simulation planning.

Furthermore, while helping behaviors by a robot are relatively easier to handle, competitive behaviors can have additional complications. While in our example from Sect. 9.1, we considered the case when *H* simulates *W*, but as seen from Fig. 6, simultaneously *W* can simulate *H* as well. Such bilateral simulations can lead to interesting behaviors, especially in competitive scenarios. In our example, suppose robot *H's* goal is to not help robot *W* but instead to impede it. Then *H* would probably decide to move one of the manipulable objects into *path A* (optimally between the two blue walls where the corridor is narrow). However, if *W* can also simulate *H's* behavior, then *W* can predict this impedance and hence, avoid *path A* and choose *path C* instead. Note that such counter simulations can go on, where *H* can predict that *W* will predict its (*H's*) impedance and choose *path C*, and thus plan to impede *W* in a different way. Bilateral simulations can thus lead to infinite inference loops. This is the typical scenario, for example in chess, where players have to predict each other's plans over multiple moves. Addition of "W's Observation of H's behavior" (dashed line in Fig. 6) during robot-robot interactions should thus be done with additional constraints.

On the other hand simulation loops cannot be avoided in human-robot interactions because, as discussed in the earlier part of this chapter, humans always simulate their interacting agent's behaviors. Therefore, the authors believe that in order to achieve intelligent interactions, robots require to acquire at least the basic mathematical/logical understanding of the psychology and neuroscience behind human interaction behaviors. The robots then need to follow these same behaviors themselves. This will allow the robots to simulate and *understand* humans when they interact with them, and enable the humans to simulate and *understand* (or at least enable them to learn quickly to understand) the robots during the interaction. Providing humans with the ability to understand robots is crucial because like an old Indian saying goes, "what we (humans) don't understand, we fear", and are uncomfortable with.

10 Conclusion

Movement notations, like from dance are essential as they provide us with essential tools to characterize and quantify movements and describe them with concise notations. In this chapter we looked at the problem of utilizing the characterized actions during interactions for understanding the actions. We started this chapter with an introduction of action understanding by humans and presented a brief review of works in philosophy, psychology and neuroscience that evaluated the mechanisms by which humans perform action understanding. We then motivated the challenges faced by robots while interacting with humans, and exhibited how the ideas from human studies, especially movement simulation can go a long way into solving these challenges. In summary what we propose is: By considering another agent to be similar to itself, a robot can utilize movement simulation in order to formulate the other agent's task, use its planners to solve the task, assume that the other agent would do the same, and then estimate which constraints impede the other agent's task the most. The agent can then plan to reduce or increase these impediments as per his required role while correcting for the estimation errors with actual observations.

The ideas we present are still conceptual where we left out many practical but essential issues like viewpoint transformations (addressed extensively in imitation studies) and especially, mechanisms to enable corrections of estimation errors. Furthermore, it is essential to develop a mechanism to integrate action understanding by movement simulation with associative learning-based understanding processes to enable a learning behavior in robots. Dance notations can play a major role in associative learning-based understanding by enabling efficient storage, characterization and retrieval of action experiences. Though some mechanisms to enable these are already available in literature, these topics need further research in the future by also incorporating knowledge of the human behavior, an issue which in itself is still far from being complete. Integrated research in robotics and neuroscience is thus required to improve future robot-human interactions.

We conclude the chapter with another interesting difference between interaction planning and the popular individual robot planning. From our discussions through this chapter, it is obvious that helping and impeding behaviors involve a larger computational work load than the task itself, because the helping agent has to first simulate the actions of the other agent inorder to understand the other agent, then plan his own task (though this may be done at a coarser resolution for some tasks), and then plan the help. This is consistent with observations in nature where the more *intelligent* animals show more complex interaction behaviors. On the other hand, it also means that agents should ideally utilize multiple plans for one task, a simpler plan for when they expect help and a complex (in terms of uncertainty and computation time) and even multiple plans (with switching) when they expect to be impeded. Thus when it comes to interaction scenarios, the optimal plan may not be a single fast solution, as is the norm with motion planning for individual robots.

Acknowledgements We thank Dr. Adrien Escande for discussions regarding our proposed concept for robot planning. This work was partially supported by JSPS KAKENHI Grant #13380602 (to G.G.) and #26120003 (to T.I.).

References

1. G. Ganesh, A. Albu Schaffer, M. Haruno, M. Kawato, E. Burdet, Biomimetic motor behavior for simultaneous adaptation of force, impedance and trajectory in interaction tasks, in *2010 IEEE International Conference on Robotics and Automation* (*ICRA*), 3–7 May 2010, pp. 2705–2711. doi:10.1109/ROBOT.2010.5509994
2. G. Ganesh, N. Jarrasse, S. Haddadin, A. Albu-Schaeffer, E. Burdet, A versatile biomimetic controller for contact tooling and haptic exploration, in *2012 IEEE International Conference on Robotics and Automation* (*ICRA*), 2012, pp. 3329–3334

3. C. Yang, G. Ganesh, S. Haddadin, S. Parusel, A. Albu-Schaeffer, E. Burdet, Human-like adaptation of force and impedance in stable and unstable interactions. Robot IEEE Trans. **27** (5), 918–930 (2011)
4. C.D. Frith, U. Frith, Interacting minds—a biological basis. Science **286**(5445), 1692–1695 (1999)
5. W. James, *Principles of Psychology* (Holt, New York, 1890)
6. T. Lipps, Die einfühlung. In Leitfaden der psychologie [Guide to psychology] Verlag von Wilhelm Engelmann, Leipzig (1903)
7. E. Stein, in *On the Problem of Empathy* (Martinus Nijhoff, The Hague, the Netherland, 1964)
8. M. Merleau-Ponty, in *Phenomenology of Perception* (trans: Smith C). (Routledge & K. Paul, London; Humanities Press, New York, 1962)
9. G.H Mead, *Mind, Self, and Society: From the Standpoint of a Social Behaviorist*, vol. 1. (University of Chicago Press, 2009)
10. L. Cardinali, F. Frassinetti, C. Brozzoli, C. Urquizar, A.C. Roy, A. Farnè, Tool-use induces morphological updating of the body schema. Curr. Biol. **19**(12), R478–R479 (2009)
11. G. Ganesh, T. Yoshioka, R. Osu, T. Ikegami, Immediate tool incorporation processes determine human motor planning with tools. Nat. Commun. **5** (2014)
12. M. Botvinick, J. Cohen, Rubber hands 'feel' touch that eyes see. Nature **391**(6669), 756 (1998)
13. H.H. Ehrsson, B. Rosen, A. Stockselius, C. Ragno, P. Kohler, G. Lundborg, Upper limb amputees can be induced to experience a rubber hand as their own. Brain **131**(Pt 12), 3443–3452 (2008)
14. H.H. Ehrsson, C. Spence, R.E. Passingham, That's my hand! Activity in premotor cortex reflects feeling of ownership of a limb. Science **305**(5685), 875–877 (2004)
15. C. Frith, in *Making Up the Mind: How the Brain Creates Our Mental World* (Wiley-Blackwell, Hoboken, 2007)
16. H.H. Ehrsson, The experimental induction of out-of-body experiences. Science **317**(5841), 1048 (2007)
17. F. de Vignemont, P. Fourneret, The sense of agency: a philosophical and empirical review of the "Who" system. Conscious. Cogn. **13**(1), 1–19 (2004)
18. W. Prinz, Perception and action planning. Eur. J. Cogn. Psychol. **9**, 129–154 (1997)
19. G. di Pellegrino, L. Fadiga, L. Fogassi, V. Gallese, G. Rizzolatti, Understanding motor events: a neurophysiological study. Exp. Brain Res. **91**(1), 176–180 (1992)
20. L. Fogassi, P.F. Ferrari, B. Gesierich, S. Rozzi, F. Chersi, G. Rizzolatti, Parietal lobe: from action organization to intention understanding. Science **308**(5722), 662–667 (2005)
21. S. Rozzi, P.F. Ferrari, L. Bonini, G. Rizzolatti, L. Fogassi, Functional organization of inferior parietal lobule convexity in the macaque monkey: electrophysiological characterization of motor, sensory and mirror responses and their correlation with cytoarchitectonic areas. Eur. J. Neurosci. **28**(8), 1569–1588 (2008)
22. G. Rizzolatti, L. Craighero, The mirror-neuron system. Annu. Rev. Neurosci. **27**, 169–192 (2004)
23. N. Nishitani, R. Hari, Temporal dynamics of cortical representation for action. Proc. Natl. Acad. Sci. USA. **97**(2), 913–918 (2000)
24. N. Nishitani, S. Avikainen, R. Hari, Abnormal imitation-related cortical activation sequences in Asperger's syndrome. Ann. Neurol. **55**(4), 558–562 (2004)
25. C. Press, J. Cook, S.J. Blakemore, J. Kilner, Dynamic modulation of human motor activity when observing actions. J. Neurosci. **31**(8), 2792–2800 (2011)
26. J.M. Kilner, A. Neal, N. Weiskopf, K.J. Friston, C.D. Frith, Evidence of mirror neurons in human inferior frontal Gyrus. J. Neurosci. **29**(32), 10153–10159 (2009)
27. T.T. Chong, R. Cunnington, M.A. Williams, N. Kanwisher, J.B. Mattingley, fMRI adaptation reveals mirror neurons in human inferior parietal cortex. Curr. Biol. **18**(20), 1576–1580 (2008)

28. V. Gazzola, C. Keysers, The observation and execution of actions share motor and somatosensory voxels in all tested subjects: single-subject analyses of unsmoothed fMRI data. Cereb. Cortex **19**(6), 1239–1255 (2009)
29. L. Fadiga, L. Fogassi, G. Pavesi, G. Rizzolatti, Motor facilitation during action observation: a magnetic stimulation study. J. Neurophysiol. **73**(6), 2608–2611 (1995)
30. M. Romani, P. Cesari, C. Urgesi, S. Facchini, S.M. Aglioti, Motor facilitation of the human cortico-spinal system during observation of bio-mechanically impossible movements. Neuroimage **26**(3), 755–763 (2005)
31. R. Mukamel, A.D. Ekstrom, J. Kaplan, M. Iacoboni, I. Fried, Single-neuron responses in humans during execution and observation of actions. Curr. Biol. **20**(8), 750–756 (2010)
32. B. Calvo-Merino, J. Grezes, D.E. Glaser, R.E. Passingham, P. Haggard, Seeing or doing? Influence of visual and motor familiarity in action observation. Curr. Biol. **16**(19), 1905–1910 (2006)
33. B. Calvo-Merino, D.E. Glaser, J. Grezes, R.E. Passingham, P. Haggard, Action observation and acquired motor skills: an fMRI study with expert dancers. Cereb. Cortex **15**(8), 1243–1249 (2005)
34. E.S. Cross, A.F.D.C. Hamilton, S.T. Grafton, Building a motor simulation de novo: observation of dance by dancers. Neuroimage **31**(3), 1257–1267 (2006)
35. S.M. Aglioti, P. Cesari, M. Romani, C. Urgesi, Action anticipation and motor resonance in elite basketball players. Nat. Neurosci. **11**(9), 1109–1116 (2008)
36. G. Hickok, in *The Myth of Mirror Neurons: The Real Neuroscience of Communication and Cognition*. (W W Norton & Co Inc, 2014)
37. G. Hickok, Eight problems for the mirror neuron theory of action understanding in monkeys and humans. J. Cogn. Neurosci. **21**(7), 1229–1243 (2009)
38. G. Csibra, in *Sensorimotor Foundations of Higher Cognition. Attention and Performance XII*. (Oxford University Press, Oxford, 2007)
39. I. Dinstein, C. Thomas, M. Behrmann, D.J. Heeger, A mirror up to nature. Curr. Biol. **18**(1), R13–R18 (2008)
40. G. Rizzolatti, L. Fogassi, V. Gallese, Neurophysiological mechanisms underlying the understanding and imitation of action. Nat. Rev. Neurosci. **2**(9), 661–670 (2001)
41. G. Rizzolatti, M.A. Arbib, Language within our grasp. Trends Neurosci. **21**(5), 188–194 (1998)
42. G. Rizzolatti, C. Sinigaglia, The functional role of the parieto-frontal mirror circuit: interpretations and misinterpretations. Nat. Rev. Neurosci. **11**(4), 264–274 (2010)
43. M.A. Umilta, E. Kohler, V. Gallese, L. Fogassi, L. Fadiga, C. Keysers, G. Rizzolatti, I know what you are doing. A neurophysiological study. Neuron **31**(1), 155–165 (2001)
44. E. Kohler, C. Keysers, M.A. Umilta, L. Fogassi, V. Gallese, G. Rizzolatti, Hearing sounds, understanding actions: action representation in mirror neurons. Science **297**(5582), 846–848 (2002)
45. M. Iacoboni, I. Molnar-Szakacs, V. Gallese, G. Buccino, J.C. Mazziotta, G. Rizzolatti, Grasping the intentions of others with one's own mirror neuron system. PLoS Biol. **3**(3), e79 (2005)
46. A.F. Hamilton, S.T. Grafton, Goal representation in human anterior intraparietal sulcus. J. Neurosci. **26**(4), 1133–1137 (2006)
47. S. Ondobaka, F.P. de Lange, M. Wittmann, C.D. Frith, H. Bekkering, Interplay between conceptual expectations and movement predictions underlies action understanding. Cereb Cortex, (2014)
48. J. Michael, K. Sandberg, J. Skewes, T. Wolf, J. Blicher, M. Overgaard, C.D. Frith, Continuous theta-burst stimulation demonstrates a causal role of premotor homunculus in action understanding. Psychol. Sci. (2014)
49. E. Tidoni, S. Borgomaneri, G. di Pellegrino, A. Avenanti, Action simulation plays a critical role in deceptive action recognition. J. Neurosci. **33**(2), 611–623 (2013)
50. Y. Kanakogi, S. Itakura, Developmental correspondence between action prediction and motor ability in early infancy. Nat. Commun. **2**, 341 (2011)

51. T. Falck-Ytter, G. Gredeback, C. von Hofsten, Infants predict other people's action goals. Nat. Neurosci. **9**(7), 878–879 (2006)
52. M.J. Pickering, S. Garrod, An integrated theory of language production and comprehension. Behav. Brain Sci. **36**(4), 329–347 (2013)
53. A. Caramazza, S. Anzellotti, L. Strnad, A. Lingnau, Embodied cognition and mirror neurons: a critical assessment. Annu. Rev. Neurosci. **37**, 1–15 (2014)
54. G. Buccino, F. Lui, N. Canessa, I. Patteri, G. Lagravinese, F. Benuzzi, C.A. Porro, G. Rizzolatti, Neural circuits involved in the recognition of actions performed by nonconspecifics: an FMRI study. J. Cogn. Neurosci. **16**(1), 114–126 (2004)
55. C. Keysers, The myth of mirror neurons the real neuroscience of communication and cognition. Science **347**(6219), 239 (2015)
56. M. Pazzaglia, N. Smania, E. Corato, S.M. Aglioti, Neural underpinnings of gesture discrimination in patients with limb apraxia. J. Neurosci. **28**(12), 3030–3041 (2008)
57. T. Ikegami, G. Ganesh, Watching novice action degrades expert motor performance: causation between action production and outcome prediction of observed actions by humans. Sci. Rep. **4**, 6989 (2014)
58. J.R. Flanagan, R.S. Johansson, Action plans used in action observation. Nature **424**(6950), 769–771 (2003)
59. J.M. Kilner, More than one pathway to action understanding. Trends Cogn. Sci. **15**(8), 352–357 (2011)
60. J.M. Kilner, K.J. Friston, C.D. Frith, Predictive coding: an account of the mirror neuron system. Cogn. Process. **8**(3), 159–166 (2007)
61. H.T. van Schie, R.B. Mars, M.G. Coles, H. Bekkering, Modulation of activity in medial frontal and motor cortices during error observation. Nat. Neurosci. **7**(5), 549–554 (2004)
62. A. Keisler, R. Shadmehr, A shared resource between declarative memory and motor memory. J. Neurosci. **30**(44), 14817–14823 (2010)
63. S.E. Criscimagna-Hemminger, O. Donchin, M.S. Gazzaniga, R. Shadmehr, Learned dynamics of reaching movements generalize from dominant to nondominant arm. J. Neurophysiol. **89**(1), 168–176 (2003)
64. G. Ganesh, M. Haruno, M. Kawato, E. Burdet, Motor memory and local minimization of error and effort, not global optimization, determine motor behavior. J. Neurophysiol. **104**(1), 382–390 (2010)
65. I.S. Howard, J.N. Ingram, D.M. Wolpert, Separate representations of dynamics in rhythmic and discrete movements: evidence from motor learning. J. Neurophysiol. **105**(4), 1722–1731 (2011)
66. T. Ikegami, M. Hirashima, G. Taga, D. Nozaki, Asymmetric transfer of visuomotor learning between discrete and rhythmic movements. J. Neurosci. **30**(12), 4515–4521 (2010)
67. J. Kodl, G. Ganesh, E. Burdet, The CNS stochastically selects motor plan utilizing extrinsic and intrinsic representations. PLoS ONE **6**(9), e24229 (2011)
68. D. Nozaki, I. Kurtzer, S.H. Scott, Limited transfer of learning between unimanual and bimanual skills within the same limb. Nat. Neurosci. **9**(11), 1364–1366 (2006)
69. D. Maslovat, N.J. Hodges, O.E. Krigolson, T.C. Handy, Observational practice benefits are limited to perceptual improvements in the acquisition of a novel coordination skill. Exp. Brain Res. **204**(1), 119–130 (2010)
70. T.L. Chartrand, J.A. Bargh, The chameleon effect: the perception-behavior link and social interaction. J. Pers. Soc. Psychol. **76**(6), 893–910 (1999)
71. M. van Elk, R. Bousardt, H. Bekkering, H.T. van Schie, Using goal- and grip-related information for understanding the correctness of other's actions: an ERP study. PLoS ONE **7**(5), e36450 (2012)
72. K. Yoshida, N. Saito, A. Iriki, M. Isoda, Social error monitoring in macaque frontal cortex. Nat. Neurosci. **15**(9), 1307–1312 (2012)
73. J.J. Gibson, *The Ecological Approach to Visual Perception* (Houghton Mifflin, Boston, 1979)

74. T.E. Horton, A. Chakraborty, R.S. Amant, Affordances for robots: a brief survey. Avant: J. Philos.-Interdisc. Vanguard, (2012)
75. P. Jacob, M. Jeannerod, The motor theory of social cognition: a critique. Trends Cogn. Sci. **9** (1), 21–25 (2005)
76. K. Friston, J. Mattout, J. Kilner, Action understanding and active inference. Biol. Cybern. **104**(1–2), 137–160 (2011)
77. K.J. Friston, The free-energy principle: a unified brain theory? Nat. Rev. Neurosci. **11**(2), 127–138 (2010)
78. D.M. Wolpert, M. Kawato, Multiple paired forward and inverse models for motor control. Neural Netw. **11**(7–8), 1317–1329 (1998)
79. D.M. Wolpert, R.C. Miall, M. Kawato, Internal models in the cerebellum. Trends Cogn. Sci. **2**(9), 338–347 (1998)
80. R. Shadmehr, S.P. Wise, *The Computational Neurobiology of Reaching and Pointing* (The MIT Press, Cambridge, Massachusetts, 2005)
81. M.S. Christensen, J. Lundbye-Jensen, S.S. Geertsen, T.H. Petersen, O.B. Paulson, J.B. Nielsen, Premotor cortex modulates somatosensory cortex during voluntary movements without proprioceptive feedback. Nat. Neurosci. **10**(4), 417–419 (2007)
82. D.M. Wolpert, J.R. Flanagan, Motor prediction. Curr. Biol. **11**(18), R729–R732 (2001)
83. R.C. Miall, L.O.D. Christensen, O. Cain, J. Stanley, Disruption of state estimation in the human lateral cerebellum. PLoS Biol. **5**(11), 2733–2744 (2007)
84. R.C. Miall, D. King, State estimation in the Cerebellum. Cerebellum **7**(4), 572–576 (2008)
85. M. Desmurget, S. Grafton, Forward modeling allows feedback control for fast reaching movements. Trends Cogns Sci. **4**(11), 423–431 (2000)
86. E. Todorov, Optimality principles in sensorimotor control. Nat. Neurosci. **7**(9), 907–915 (2004)
87. E. Todorov, M.I. Jordan, Optimal feedback control as a theory of motor coordination. Nat. Neurosci. **5**(11), 1226–1235 (2002)
88. Y.W. Tseng, J. Diedrichsen, J.W. Krakauer, R. Shadmehr, A.J. Bastian, Sensory prediction errors drive cerebellum-dependent adaptation of reaching. J. Neurophysiol. **98**(1), 54–62 (2007)
89. R. Shadmehr, M.A. Smith, J.W. Krakauer, Error correction, sensory prediction, and adaptation in motor control. Ann. Rev. Neurosci. **33**, 89–108 (2010)
90. M. Haruno, D.M. Wolpert, M. Kawato, Mosaic model for sensorimotor learning and control. Neural Comput. **13**(10), 2201–2220 (2001)
91. D.M. Wolpert, Z. Ghahramani, Computational principles of movement neuroscience. Nat. Neurosci. **3**(Suppl), 1212–1217 (2000)
92. M. Kawato, Internal models for motor control and trajectory planning. Curr. Opin. Neurobiol. **9**(6), 718–727 (1999)
93. S.J. Blakemore, J. Decety, From the perception of action to the understanding of intention. Nat. Rev. Neurosci. **2**(8), 561–567 (2001)
94. R. Grush, The emulation theory of representation: motor control, imagery, and perception. Behav. Brain Sci. **27**(3), 377–396; discussion 396–442 (2004)
95. M. Jeannerod, Neural simulation of action: a unifying mechanism for motor cognition. Neuroimage **14**(1 Pt 2), S103–S109 (2001)
96. D.M. Wolpert, K. Doya, M. Kawato, A unifying computational framework for motor control and social interaction. Philos. Trans. R. Soc. Lond. B Biol. Sci. **358**(1431), 593–602 (2003)
97. J. Perner, Simulation as explication of prediction-implicit knowledge: re-assessing its value for explaining the development of mental state attributions, in *Theories of Mind*, eds. by P. Carruthers, P.K. Smith (Cambridge University Press, United Kingdom, 1996)
98. R.M. Gordon, Folk psychology as simulation. Mind Lang. **1**(2), 158–171 (1986)
99. M. Haruno, D.M. Wolpert, M. Kawato, Hierarchical MOSAIC for movement generation. Int. Congr. Ser. **1250**, 575–590 (2003)
100. S. Schaal, Is imitation learning the route to humanoid robots? Trends Cogn. Sci. **3**(6), 233–242 (1999)

101. M. Lopes, F. Melo, L. Montesano, J. Santos-Victor, Abstraction levels for robotic imitation: overview and computational approaches, in *From Motor Learning to Interaction Learning in Robots. Studies in Computational Intelligence*, vol. 264 eds. by O. Sigaud, J. Peters. (Springer Berlin Heidelberg, pp. 313–355, 2010). doi:10.1007/978-3-642-05181-4_14
102. M. Lopes, J. Santos-Victor, Visual learning by imitation with motor representations. IEEE Trans. Syst. Man Cybern. Part B: Cybern. **35**(3), 438–449 (2005)
103. M. Asada, Y. Yoshikawa, K. Hosoda, Learning by observation without three-dimensional reconstruction. Intell. Auton. Syst. (IAS-6), 555–560 (2000)
104. A. Alissandrakis, C.L. Nehaniv, K. Dautenhahn, Action, state and effect metrics for robot imitation, in *The 15th IEEE International Symposium on Robot and Human Interactive Communication*, 2006. ROMAN 2006. 6–8 September 2006, pp. 232–237. doi:10.1109/ROMAN.2006.314423
105. S. Nakaoka, A. Nakazawa, K. Yokoi, H. Hirukawa, K. Ikeuchi, Generating whole body motions for a biped humanoid robot from captured human dances, in *Proceedings of IEEE International Conference on Robotics and Automation, 2003*. ICRA '03, vol. 3903. 14–19 September 2003, pp. 3905–3910. doi:10.1109/ROBOT.2003.1242196
106. W. Erlhagen, A. Mukovskiy, E. Bicho, G. Panin, C. Kiss, A. Knoll, H. van Schie, H. Bekkering, Goal-directed imitation for robots: a bio-inspired approach to action understanding and skill learning. Robot. Auton. Syst. **54**(5), 353–360 (2006)
107. W. Yang, N.Y. Chong, C.H. Kim, B.J. You, Locomotion imitation of humanoid using goal-directed self-adjusting adaptor, in *International Conference on Intelligent Robots and Systems*, 2006 IEEE/RSJ, 2006. IEEE, pp 5650–5656
108. M. Lopes, J. Santos-Victor, A developmental roadmap for learning by imitation in robots. IEEE Trans. Syst. Man Cybern. Part B: Cybern. **37**(2), 308–321 (2007)
109. K. Mochizuki, S. Nishide, H.G. Okuno, T. Ogata, Developmental human-robot imitation learning of drawing with a neuro dynamical system, in *2013 IEEE International Conference on Systems, Man, and Cybernetics* (*SMC*). 13–16 October 2013, pp. 2336–2341. doi:10.1109/SMC.2013.399
110. L. Baker, J.B. Tenenbaum, R.R. Saxe, Bayesian models of human action understanding. Adv. Neural Inf. Process. Syst. **18**, 99–105 (2006)
111. A. Boularias, B. Chaib-draa, in *Learning Probabilistic Models via Bayesian Inverse Planning. Poster at the Neural Information Processing Systems*, 2009
112. N. Ratliff, B. Ziebart, K. Peterson, J.A. Bagnell, M. Hebert, A.K. Dey, S. Srinivasa, in *Inverse Optimal Heuristic Control for Imitation Learning* (AISTATS, 2009)
113. B.D. Ziebart, A.L. Maas, J.A. Bagnell, A.K. Dey, Maximum entropy inverse reinforcement learning, in *AAAI*, pp. 1433–1438 (2008)
114. A. Billard, Y. Epars, S. Calinon, S. Schaal, G. Cheng, Discovering optimal imitation strategies. Robot. Auton. Syst. **47**(2–3), 69–77 (2004)
115. C.L. Baker, R. Saxe, J.B. Tenenbaum, Action understanding as inverse planning. Cognition **113**(3), 329–349 (2009)
116. J-C. Latombe, in *Robot Motion Planning: Edition en anglais*. (Springer, Berlin, 1991)
117. LaValle SM, in *Planning Algorithms*. (Cambridge University Press, United Kingdom, 2006)
118. S. Dalibard, A. El Khoury, F. Lamiraux, A. Nakhaei, M. Taïx, J-P. Laumond, Dynamic walking and whole-body motion planning for humanoid robots: an integrated approach. Int. J. Robot. Res. 0278364913481250 (2013)
119. Y. Nakamura, H. Hanafusa, Inverse kinematic solutions with singularity robustness for robot manipulator control. J. Dyn. Syst. Measur. Control-Trans. Asme **108**(3), (1986)
120. B. Siciliano, J.J.E. Slotine, A general framework for managing multiple tasks in highly redundant robotic systems, in *Fifth International Conference on Advanced Robotics*, *1991. 'Robots in Unstructured Environments'*, 91 ICAR, vol. 1212. 19–22 June 1991, pp. 1211–1216. doi:10.1109/ICAR.1991.240390
121. P. Baerlocher, R. Boulic, Task-priority formulations for the kinematic control of highly redundant articulated structures, in *Proceedings of IEEE/RSJ International Conference on*

Intelligent Robots and Systems, vol. 321, 13–17 October 1998, pp. 323–329. doi:10.1109/IROS.1998.724639
122. O. Khatib, L. Sentis, J. Park, J. Warren, Whole-body dynamic behavior and control of human-like robots. Int. J. Humanoid Rob. **01**(01), 29–43 (2004)
123. F. Kanehiro, F. Lamiraux, O. Kanoun, E. Yoshida, J.-P. Laumond, A local collision avoidance method for non-strictly convex polyhedra, in *Proceedings of Robotics: Science and Systems IV,* 2008
124. L.E. Kavraki, P. Svestka, J.C. Latombe, M.H. Overmars, Probabilistic roadmaps for path planning in high-dimensional configuration spaces. IEEE Trans. Robot. Autom. **12**(4), 566–580 (1996)
125. J.J. Kuffner, S.M. LaValle, Space-filling trees: a new perspective on incremental search for motion planning, in *IEEE/RSJ International Conference on Intelligent Robots and Systems* (*IROS*). IEEE, pp. 2199–2206 (2011)
126. J.C. Barca, Y.A. Sekercioglu, Swarm robotics reviewed. Robotica **31**(03), 345–359 (2013)
127. T. Abukhalil, M. Patil, T. Sobh, Survey on decentralized modular swarm robots and control interfaces. Int. J. Eng. (IJE) **7**(2), 44 (2013)
128. G. Ganesh, A. Takagi, R. Osu, T. Yoshioka, M. Kawato, E. Burdet, Two is better than one: physical interactions improve motor performance in humans. Sci. Rep. **4**, (2014)
129. C. Frith, Keynote Lecture. Ninth Annual Meeting of the Organisation for Human Brain Mapping. Neural Hermeneutics: how brains interpret minds (2003)
130. D.M. Wolpert, J. Diedrichsen, J.R. Flanagan, Principles of sensorimotor learning. Nat. Rev. Neurosci. **12**(12), 739–751 (2011)
131. G. Ganesh, E. Burdet, Motor planning explains human behaviour in tasks with multiple solutions. Robot. Auton. Syst. **61**(4), 362–368 (2013)

Challenges for the Animation of Expressive Virtual Characters: The Standpoint of Sign Language and Theatrical Gestures

Sylvie Gibet, Pamela Carreno-Medrano and Pierre-Francois Marteau

Abstract Designing and controlling virtual characters endowed with expressive gestures requires the modeling of multiple processes, involving high-level abstract representations to low-level sensorimotor models. An expressive gesture is here defined as a meaningful bodily motion which intrinsically associates sense, style, and expressiveness. The main challenges rely both on the capability to produce a large spectrum of parametrized actions executed with some variability in various situations, and on the biological plausibility of the motion of the virtual characters. The goals of the paper are twofold. First we review the different formalisms used to describe expressive gestures, from notations to computational languages. Secondly we identify and discuss remaining challenges in the generation of expressive virtual characters. The different models and formalisms are illustrated more particularly for theatrical and sign language gestures.

1 Introduction

Performing skilled gestures and movements, eventually in interaction with users and the environment, requires a thorough understanding of the different levels of representation that underlay their production, from the construction of sense to the elaboration of motor programs that prefigure motion performances. This is even truer for meaningful and expressive gestures which involve high level semiotic and cognitive representations, and require rapidity, accuracy, and physical engagement with the environment.

During the past decades, methods and approaches describing and modeling realistic human movements have been largely investigated by the research community in many areas such as motion analysis, recognition and synthesis, and

S. Gibet (✉) · P. Carreno-Medrano · P.-F. Marteau
IRISA, University of Bretagne Sud, Campus de Tohannic, rue Yves Mainguy,
56017 Vannes, France
e-mail: sylvie.gibet@univ-ubs.fr

J.-P. Laumond and N. Abe (eds.), *Dance Notations and Robot Motion*,
Springer Tracts in Advanced Robotics 111,
DOI 10.1007/978-3-319-25739-6_8

contributed to a wide range of applications using virtual characters or humanoid robots.

However, if technologies of interactive embodied conversational agents are now available to researchers, there are still many open challenges concerning the animation of virtual characters endowed with expressive behavior. As pointed out by Thiebaux et al., virtual humans must be *believable*, *interpretable*, and *responsive* [1]. A believable character can be defined as "one that provides the illusion of life" [2]. This can be characterized by the perceivable behavior, in terms of motion consistency and quality, as well as the avatar's appearance. The interpretability, that we will call here comprehension, concerns both the user's ability to understand the message conveyed by the avatar's gestures, and the expressive information encoded in the movement. The responsiveness, which involves the property of reactivity, is related to the ability of the virtual character to respond to events from the environment, and in particular it makes possible the interaction with the user. To go further, researchers aim at designing compelling characters capable of creating a more intuitive and engaging interaction with a user.

In this paper, beyond the properties of believability, comprehension, and responsiveness, we also aim at highlighting the main difficulties in characterizing, modeling and producing expressive behaviors driven by an underlying semantics. We review the existing formalisms and concepts used to describe expressive gestures, from notations to computational languages, how this specification may influence the produced movements, and discuss the remaining challenges for animating expressive virtual characters. We do not pretend to provide an exhaustive overview of the different technologies used to create credible virtual characters, as proposed for human-computer dialog [3], but focus more specifically on full-body movements which draw the user's attention, and express through body language some meaningful and emotional intent. The different issues will be addressed for two categories of movements (i) theatrical gestures which are demanding in the production of believable and engaging movements, and, (ii) sign language gestures which involve highly structured movements driven by constrained linguistic rules, thus pushing the comprehension to a demanding level. Both categories of movements implicitly contain a strong semantics and involve complex cognitive, linguistic and sensorimotor mechanisms, from story telling to the production of movements whose detailed components may contain significant elements perceivable by humans.

2 Requirements for Producing Expressive Gestures

Understanding the mechanisms involved in the production of meaningful and expressive gestures implies strong directions for future research. Some essential and not yet accomplished characteristics that make the production of such gestures highly complex are exposed below, such as multi-modality, spatial content, coordination/synchronization rules, and expressiveness. These requirements are highlighted and illustrated in the context of the execution of sign language gestures and theatrical movements.

2.1 Multimodal Components

Gestures are not restricted to conveying meaning solely through body movements; instead, they require the simultaneous use of body, hands' movements, facial mimics, gaze direction, and speech. In sign languages (SL), we generally separate the manual components which include hand configuration, orientation, and placement or movement, expressed in the signing space (the physical three-dimensional space in which the signs are performed), from non-manual components consisting of the posture of the upper torso, head orientation, facial expression, and gaze direction. For example, eye gaze can be used to recall a particular object in the signing space; it can also be necessary to the comprehension of a sign, as in the sign DESSINER(v) corresponding to the action of drawing, and for which the eyes follow the motion of the fingers as in drawing.

In dance, the eye gaze may be used for balance purpose, or in a magical trick for distracting the attention of the spectator. In theatrical gestures (TG), a role has been attributed to each non-manual component according to Delsarte's movement theory. For instance, the torso is considered as the main channel of emotional content while arms act as thermometers indicating how expressive a movement can be [4].

In SL, facial mimics may serve as adjectives (e.g., inflated cheeks make an object large or cumbersome, while squinted eyes make it thin) or indicate whether the sentence is a question (raised eyebrows) or a negation (frowning). It is therefore very important to preserve this information during facial animation. Both in TG and SL, facial mimics convey the emotion and the state of mind of the signer/actor, which is useful for the comprehension of the various situations, but also for the credibility and the correctness of the movement as well as the expressiveness intent.

In theater, speech and its paralinguistic characteristics such as pitch, loudness, tempo, among others, "[…] constitute one of the most ancient objects of the actor's art." [5]. Actors control the flow of information and segment discourse in order to increase the engagement and comprehension levels of the audience. They also use speech as a means of strengthening the information about the motivations and emotions behind the characters they personify [5].

2.2 Spatial Content

TG or SL use by nature spatial mechanisms, in particular for strongly iconic gestures, i.e. gestures that describe the shape of an object or an entity, or mime a situation in space. Both SL and TG execute movements involving a part or several parts of the body, for example raising a shoulder or nodding and shaking head.

In SL, spatiality is intrinsically linked to meaning and is mostly expressed through hand movement trajectories, or hand shapes (called configurations). Thus, spatial mechanisms, classically used in LSF (French Sign Language) are depicting or directional verbs which mimic spatial movements, as well as size-and-shape configurations which are static spatial descriptions. In TG, similarly to SL, the orientation of gesture in space, the starting and ending point in space of a movement, all impact the meaning of a gesture [6].

Movement trajectories. In SL, indicating and depicting verbs require the signer to manipulate targets in the signing space by effecting pointing-like movements towards these targets. They include such signs as the LSF sign DONNER (v., give), in which the hand moves from the giver to the receiver. Depending on the intended subject and object, the initial and final placements of the hand, as well as its orientation vary greatly within the signing space; these placements may have syntactic meaning (subject, object, pronoun, etc.). Note that for such categories of signs, the target can be located on a specific part of the body. Other more accurate spatial mechanisms involve hand trajectories within signs, which are not only transitions in space between two key positions, but take the shape of a line, an arc, or a more complex form such as an ellipse, a spiral, etc. For example the sign AIMER (v., like) is represented by an upward arc-movement. An interesting inquiry is whether playing reversible indicating verbs backwards would be convincing to other signers, and whether altering the hand-shape of a stored depicting verb would be understood as a change in meaning. Thus, we can reverse the movement of AIMER to produce the meaning NE-PAS-AIMER (v., dislike, see Fig. 1, right). In the same way, reversing the sign DONNER (v., give) may result in the LSF sign PRENDRE (v., take, see Fig. 1, left). These mechanisms can lead to a gestural language where the trajectory is spatially coded [7].

In TG, it has been suggested that the trajectory a movement describes in space can have an important expressive content as well as capture the overall tendency of the motion [4]. Changes in motion trajectory can also be used to emphasize and accentuate a movement in the eyes of the audience.

Shapes. Moreover, the depicting verbs can be performed with a generic hand-shape, or with a hand-shape that indicates the object has a cylindrical shape. Using hand-shapes, we also find signs in which one hand acts as the dominated hand, while the other is the dominant one. For example, in the case of the LSF sign

Fig. 1 LSF signs LIKE/LIKE-NOT (*left*), and GIVE/TAKE (*right*): the hand trajectories are reversed

AVION (plane), the flat dominated hand is representing the runway, the other one the plane taking off. Finally, the signing space may be decomposed into different relational spaces: the discourse space, in which the entities of the discourse are represented and located. The topographic space, which expresses the spatial relationships between the entities, and may use embedded layouts of spatial descriptions. In TG, meanings and emotions are encoded into specific body postures that may be exaggerated in shape and duration to ensure the understanding of the interlocutor. The shape taken by the body concerns all contours of the bodies made in space. Shapes can be of three types: lines, curves, or combination of both performed in the three main planes (frontal, sagittal, or horizontal). They are stationary or moving through space and depict information about how a person interacts with her surroundings [8].

Precision. In SL, comprehension of signs requires accuracy in their formation. Some hand-shapes differ only by the position of one finger or by whether or not it contacts another part of the hand or the body. In addition, the degree of openness of the fingers can be the sole differentiating factor between signs. This calls for notable accuracy in the motion capture and data animation processes. In TG, precision and simplicity in an actor's motion are fundamental for a clear and successful communication with the audience. TG are based on a combination of simplification (to bring focus to a particular element by eliminating all possible superfluous and ambiguous movement) and exaggeration (after simplifying a movement, emphasize its meaning by exaggerating it) principles.

2.3 Coordination/Synchronization Rules

In order to increase the controllability of the movements, it appears essential to be able to manipulate the movements at a finer grain than the postures themselves. This requires the precise decomposition of the body along channels that are significant to the manipulated motion type, and the definition of coordination/synchronization rules. Both in SL and TG, altering the spatio-temporal properties of the movements may deeply modify the meaning of the gestures. We give hereinafter different temporal aspects and constraints that characterize the execution of gestures.

For SL, the question of timing and dynamics of gesture is crucial. In fact, three elements are of interest for these gestures. Firstly, in SL, the kinematics characterizing the transition movements and the stroke conveying a specific meaning shows specific profiles that should be respected when synthesizing such gestures. Secondly, spatio-temporal synchronization rules between different parts of the body is a major component. In particular, phonetic studies have shown structural patterns with regular temporal invariants [9], such as the hand configuration target which is systematically reached before the hand movement begins [10], or the motion of the two hands which are very often synchronized, the dominant hand slightly preceding the non dominant hand. We may also observe some timing invariants between eye gaze and head movements, or eye-gaze and hand configuration. Thirdly, the dynamics of the gesture (acceleration profile along time) can be used to distinguish two meanings. An example is the difference between the LSF signs JOUER(v) (to play), and DÉTENDU (relaxed), which have the same hands configurations, the same trajectories in space, but different dynamics. Let us finally note that the dynamics of contacts between the hand and the body (gently touching or striking) is particularly relevant.

In TG, temporal characteristics as tempo, duration and repetition are highlighted and can be used as a language of gesture for creating theatrical compositions [8]. Additionally, it is important to spend the correct amount of time on each movement. Otherwise, the audience will not be able to follow and interpret the inner intentions and motivations of the character personified by the actor. It is also possible to indicate a character's current state of mind by modifying the timing of its actions.

3 Previous Work

Many studies have addressed the problem of describing, categorizing, and formalizing computational models for generating movements. We summarize below some knowledge from experts in theatrical and sign language movements in terms of gesture descriptions and notations. We then derive the main trends used to animate expressive and meaningful virtual characters, and discuss the good points as well the main drawbacks in meeting the previous requirements.

3.1 Gesture Descriptions and Notations

Early work in linguistics has attempted to describe and categorize movements and gestures. For gestures conveying a specific meaning, called *semiotic* gestures, taxonomies have been proposed. They define *semantic* categories, i.e. semantic classes that can be discriminated and characterized by verbal labels. Kendon [11] is the first author to propose a typology of semiotic acts, making the hypothesis of a continuum between speech utterances and information conveyed by gestures. McNeill extends this typology with a theory gathering the two forms of expression, speech and action [12]. In these studies, both modalities are closely related, since they share a common cognitive representation. Furthermore, Kendon and McNeill [11, 12] have proposed a temporal structure of gestures, above all for co-verbal gestures. This structure can be described in terms of phases, phrases, and units. It has been extended by Kita [13] who introduced the different phases (Preparation, Stroke, and Retraction) composing each significant unit, and used in the context of SL generation [14].

In order to memorize and transcribe the gestures and their structural organization, movement notations and coding systems have also been developed. These systems generally aim at describing in an exhaustive and compact way labeled elements whose structure relies on a predefined vocabulary depending on the studied movement and context.

Laban Movement Analysis (LMA). Among these structural descriptions, the Laban Movement Analysis (LMA) theory initially defined for dance choreography identifies semantic components that describe the structural, geometric and dynamic properties of human motion [15, 16]. This theory comprises four major components: Body, Space, Effort and Shape. Body and Space components describe how the human body moves, either within the body or in relation with the 3D space surrounding the body. Shape component describes the shape morphology of the body during the motion, whereas Effort component focuses on the qualitative aspects of the movement in terms of dynamics, energy and intent. LMA has been largely used in computer animation [17–19], motion segmentation [20], gesture recognition and affect analysis [21, 22].

Eshkol-Wachman notation system. Although initially developed for dance, it was also intended to notate and analyze any possible movement in space in a rather mathematical way [23]. The moving body is treated as a system of articulated axes in which each axis corresponds to a line segment of constant length connecting either two joints or a joint and a free extremity. The path described by each axis's end is parameterized using spherical-like coordinates. The notation system also describes three types of movements: rotatory, plane and conical and describes the degree of interdependence between limbs as *light* or *heavy*. All limbs are thus divided into relative classes: every limb is *heavy* relatively to any limb that it carries while moving, and *light* relatively to any limb by which it is being carried. This system has been used in a wide variety of fields like sports [24], sign language [25], medicine [26], etc.

Delsarte notation system. It is a notation system based on Delsarte's (a French musician and actor) methodical observations of the human body and its interactions with others. Through his notation system, Delsarte described the relationship between meaning and motion, and how attitude and personality are conveyed by body parts and gestures [27]. Motions are classified into three categories according to the direction of movement: *eccentric*, motion away from the body center and having a relation to the exterior world; *concentric*, motions toward the body center and having relation to the interior; and *normal*, balanced motions moderating between concentric and eccentric motions. The body is divided into body zones, each zone having nine possible poses i.e., all combinations of the three types of motion. For each pose and each zone a meaning is attributed. Delsarte identified three orders of movement: *oppositions*, *parallelisms*, and *successions* as well as nine laws of motion (attitude, force, motion, sequence, direction, form, velocity, reaction and extension) that further modify the meaning of each movement. Delsarte system has been already used for the generation of virtual agents motions [6, 28].

Sign Language descriptions. The notion of decomposing signs into various components is not new to the linguistic community. In 1960, William Stokoe started his system of *Tab* (location), *Dez* (handshape), and *Sig* (movement) specifiers that were to describe any sign [29]. Since then, other linguists have expanded on Stokoe's decompositional system, keeping the location, hand-shape, and placement parameters, and introducing wrist orientation, syllabic patterning, etc. [9, 30]. All systems allow for multiple configurations during the course of a single sign. In her 1981 work, Sutton presents the SignWriting model, using pictographic symbols placed spatially to represent sign components including their associated facial expressions. HamNoSys [31], another pictographic notation system, is a Stokoe-based phonetic notation system, using a linear presentation of custom symbols with diacritics. However, this system provides no timing information in the form of sign segments, and thus makes an analysis of sign timing rather difficult. The system proposed by Liddell and Johnson introduces timing segments that divide signs into sequential parts in addition to the simultaneous parts that Stokoe had previously identified. Hand configuration, points of contact, facing, and orientation are described as static articulatory postures; movements allow for spatial and temporal progression between postures. Following Liddell and Johnson's phonetic patterns, and the grammatical decomposition proposed by Johnston and de Beuzeville [32], Duarte et al. have developed their own annotation scheme which is used for the synthesis of signing utterances in French sign language [33].

3.2 Animation of Virtual Characters

The modeling of human-like behavior leads to an intelligent virtual agent generally considered as deliberative, since it has the capability of decision, and reactive in the sense it can react to events. This requires the integration of both cognitive and reactive aspects, based on the will and intention of the agent, as well as on perceptuo-motor processes occurring during the motor performances. Different trends in the research on cognitive architectures have recently emerged, highlighting the role of memory and learning in the design of intelligent systems that have similar capabilities to those of humans. Two surveys review various paradigms of cognition [34], and various architectures among *symbolic*, *emergent*, and *hybrid* models [35]. However, there are very few cognitive architectures that are implemented and applied to the animation of virtual characters. Among these systems, the concept of Action/Perception/Decision has given rise to a programming environment for behavioral animation [36].

Many levels have been defined for behavior planning and control, and for specification languages dedicated to expressive virtual characters [3]. Two major classes of approaches can be distinguished: (i) those that specify explicit "intelligent" behaviors dedicated to embodied conversational agents, and (ii) those offering data-driven animation techniques. Some hybrid frameworks combine these two approaches to respond to the requirements stated above.

Embodied Conversational Agents (ECA). Creating ECAs requires designing high-level behavior (planning, handle communicative acts, etc.), and producing coordinated and synchronized movements of multiple parts of the body, possibly associated with speech production: upper and lower body, head/neck, hands, facial expression, eye movements, speech. Regarding high-level gesture specification, historical and current methods range from formalized scripts to dedicated gestural languages. The Behavior Expression Animation Toolkit (BEAT), as one of the first systems to describe the desired behaviors of virtual agents, uses textual input to combine gesture features for generation and synchronization with speech [37]. XML-based description languages have been developed to describe various multi-modal behaviors, some of which are dedicated to complex gesture specification [38], describe style variations in gesture and speech [39], or introduce a set of parameters to categorize expressive gestures [40]. More recently, some computational models consider the coordination and adaptation of the virtual agent with a human or with the environment in interacting situations. The models in such cases focus on rule-based approaches derived from social communicative theories [41].

To facilitate the creation of interactive agents, recent work has proposed the SAIBA architecture, in which three main stages are identified, namely the intent planner, the behavior planner, and the surface realizer [42, 43]. This software architecture is the basis for implementing various embodied characters with unified and abstract interfaces. The functional markup language (FML) is used to encode the communicative intent, whereas the behavior markup language (BML) specifies the verbal utterance and the nonverbal behaviors such as gesture or facial expression (e.g., pointing gesture, shaking hands, nodding head, etc.).

Passing from the specification of gestures to their generation has given rise to a few research work. Largely, this work aims at translating a gestural description, expressed in any of the above-mentioned formalisms, into a sequence of gestural commands that can be directly interpreted by a real-time animation engine. Most of the animation models rely on pure synthesis methods, for example by using inverse kinematics techniques (e.g., [44, 45]).

More recently, novel languages and architectures, based on the SAIBA-BML behavior language have been proposed. The SmartBody [1] is an open source modular framework which hierarchically interconnects controllers to achieve continuous motion. It employs various animation algorithms such as key-frame interpolation, motion capture or procedural animation. The real-time system EMBR introduces a new animation layer of control between the behavioral level and the procedural animation level, thus providing the animator with a more flexible and accurate interface for synthesizing nonverbal behaviors [46]. This system also incorporates into the language expressive parameters (spatial extent, temporal extent, fluidity, and power) [40]. Most of the proposed languages describe the behaviors in an explicit way, thus preventing the system's ability to respond reactively to external events, or to anticipate the movement of some body parts in complex tasks. Without offering an animation specification language, the PIAVCA architecture [47] proposes a functional abstraction of character behavior. It provides a range of motion editing filters that can be combined to achieve animations reactive to events.

The approaches using high-level specification languages coupled with interpolation or procedural animation methods allow for building complex behaviors, essentially by combining different controllers associated to different modalities. Moreover, based on psycho-linguistic rules or manual annotations, the generated movements are consistent and precise. Finally, some scripting language may take into account expressiveness in terms of semantic labels or expressive parameters.

However, building by hand complex animations by specifying key postures or targets and synchronizing the different body parts in space and time has revealed to be a tedious task. Another drawback of such methods is the lack of believability for generating motion, except for those that use motion capture controllers. In order to ease and automatize the generation of novel movement sequences, it is necessary to take into account some movement knowledge in terms of structural spatio-temporal patterns, human motion rules (such as invariant motion laws), or statistical motion properties. To summarize, the most significant benefits of the ECA related methods are the controllability and the precision of the behavior of the virtual character, but this is achieved at the expense of ease of specification and believability.

Data-driven Synthesis. Alternatively, to achieve animation of highly believable and life-like characters, data-driven methods have replaced pure synthesis methods. In this case the movements of a real user are captured with different combinations of motion capture techniques. Motion graphs allow to generate realistic, controllable motion through a database of motion capture [48]. The authors automatically construct a graph that encapsulates connections among different motion chunks in the database and then search this graph for motions that satisfy user constraints. One limitation of the approach is that the transition thresholds must be specified by hand, which may prove to be a very tedious task.

Furthermore, machine learning techniques can be used to capture style in human movements and generate new motions with variations in style or expressiveness [49–52]. In these studies authors consider a low-level definition of style, in terms of variability observed among several realizations of the same gesture. If some relevant studies rely on qualitative or quantitative annotations of motion clips (e.g., [53, 54]), or propose relevant methods to create a repertoire of expressive behaviors (e.g., [55]), very few approaches deal with both motion-captured data and their implicit semantic and expressive content. Within their framework, Stone et al. synchronize meaningfully gesture and speech by specifying the organization of characters' utterances and generating automatically the animation of the conversational character [56]. The authors rely on an annotation process that indicates the perceptually prominent moments of emphasis in speech and gesture. To animate gesturing characters, Jörg et al. develop a motion retrieval method to automatically add plausible finger motions to body motions, extracting the finger motions from a database, according to the similarity of the arm movements and the smoothness of finger motions [57]. To create natural-looking motions of characters that follow users scenarios, Safonova et al. provide a sketched-based method associated to a motion graph representation to approximatively specify the path of the character and adapt the existing motions through interpolation [58].

These approaches give satisfactory results in terms of believability, since they use postures or motion chunks selected in a pre-defined database. It still remains difficult to parameterize motion and to produce controllable and flexible behaviors. In addition, the reuse of motion data does not give the ability to generate novel movements far from existing ones. Another drawback is the lack of responsiveness of such fully data-driven approach.

4 Remaining Challenges to Animate Expressive Virtual Character

In this section only data-driven methods are considered, as they particularly meet the necessary requirement of believability of the produced animated sequences. Though these methods significantly improve the believability of the animations, there are nonetheless several remaining challenges to the reuse of motion data. The main one is the transformation and recombination of motion capture data elements in the production of behaviors that preserve the movement's sense and the emotional intent. We discuss hereafter these challenges following the requirements evoked in Sect. 2.

4.1 Constructing Resources with Meaning and Expressiveness

Data acquisition. Signs and theatrical gestures are by nature expressive and dexterous gestures, which simultaneously involve several modalities (arms, hands, body, gaze and facial expressions). Capturing accurately and synchronously all these channels with an appropriate frequency (>100 MHz) actually pushes motion capture equipment to their limits. Novel technologies such as surface capture [59], that captures simultaneously geometry and animation, are very attractive, but yet the resolution is not sufficient to capture the body and the face with an adequate precision, and only few methods exist to manipulate this complex data in order to produce new animations.

Nature of the gesture corpus. For the purpose of corpus design, several questions have to be addressed. The first one concerns the corpus definition and the compromise that exists between breadth and depth in its design. If the objective of the synthesis system is to have a lexicon that covers a broad area, including several thematic domains, then a corpus with a breadth approach would be suitable. If, instead, the goal is to have a limited vocabulary and reuse it in different sentences, then the depth approach would be best. In this case, many tokens of the same signs or actions will be preferred in the predefined vocabulary, with variations depending on the scenario context. The second question concerns the nature of the variations that should be included in the corpus for further editing and synthesis. Several levels of signs variability can be considered: we may think about incorporating multiple tokens of the same action/sign in different contexts, in order to be able to construct new sentences that take into account the spatial variations as well the co-articulation aspects. For example, a magician in theatrical gestures might want to perform his trick in different locations in space, or describe objects of different shapes and sizes. The problem is similar in SL, but with finer motion elements (manipulating short hand movements or hand configurations). The signing context can also be taken into account by capturing the same sign in varying its predecessors and successors (e.g. influence of hand shape and placement). The inclusion of such sequencing in the corpus allows for the study of co-articulation. Therefore, if the actions/signs involving such inflection processes are contained into the corpus, then the editing operations will be less complex. For example, including many depicting verbs with spatial variation will facilitate the construction of novel utterances with verb declination, without recording new signs. Another source of variation is the style of the actor/signer, or the deliberate emotional state contained in the scenarios, which lead to kinematic variations. A closely linked question concerns the acted or spontaneous nature of the produced scenarios.

4.2 High Level Language for Multichannel Editing

The choice of the computing language allowing the description of behaviors that can be interpreted by the animation controllers is still very challenging to the computer animation community, above all for communicative and expressive behaviors involving high level semantic rules. Most of the time, these behaviors concern the combination and scheduling of elementary behaviors attached to dedicated controllers such as keyframe interpolation, motion capture, or procedural animation. This approach does not consider the coordination of finer-grain motion which is necessary when dealing with communicative gestures. Using a predefined dual database, one containing the raw motion and the other annotated data, it becomes possible to build novel phrases, by selectively composing and blending pre-existing elements along temporal segments and spatial channels [60]. In this scope, it is necessary to consider all the unsolved spatial and temporal issues raised by the editing process.

Inflecting spatial variations. When dealing with motion capture data, it is very difficult to generate new movements that are not contained in the database. However, a main challenge would be to define generic and parameterized controllers that enable the generation of similar motions varying in space, for example if we want to modify the location of a gesture (up-right or down-left), or the size of an object (showing a small or big box).

Spatial coherency. Another challenge would be to combine different spatial channels with different meanings simultaneously. This coordination differs from the classical blending approaches which mix whole skeleton motions to produce new

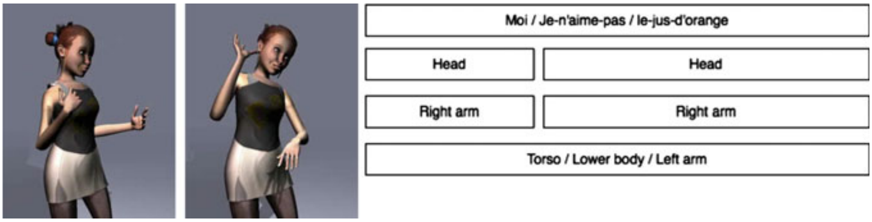


Fig. 2 Combination of three signs: moi/je-n'aime-pas/le-jus-d'orange (I don't like orange juice)

ones [53]. An example is given in Fig. 2 which illustrates the construction of the sentence: "I don't like orange juice" in LSF. Different channels are combined, by keeping the torso/lower body/left arm of one sequence, and substituting the head, facial expression and right arm movements of another sequence. In such a composition process, the spatial constraints should be preserved, in particular the sign should be executed near the corresponding body part, whatever the torso or the head orientation is. This clearly reveals that the combination process should be driven at a more abstract level, expressed by rules or constraints incorporated into the controllers.

Inflecting temporal variations. It is likely that the different motion elements have not the same duration. The subsequent problem is twofold: (i) a common timeline has to be found, eventually as the result of a combinatorial optimization, or driven by linguistic rules. Up to our knowledge though, no existing gestural language describes such temporal rules or models the synchronization of the different channels (ii) once a correct time plan has been devised, the temporal length of the motion chunks has to be adapted, while preserving the dynamics of the motions. To this end, time warping techniques can be used [61]. However, inter channels synchronizations may exist (for example between the hand and the arm motions [62]). Thus synchronization schema can be extracted from analysis, but the proper way to introduce this empirical knowledge in the synthesis process has not been explored yet.

4.3 *Dealing with Expressiveness*

Virtual characters portray their emotions through movement [63], thus as stated by Byshko [64], the perception of a virtual character has everything to do with how it moves. Unfortunately, in spite of the numerous psychological studies and computer animation and machine learning applications trying to decode and exploit the most salient features to human expressiveness, there is still no common understanding about how affect, style and intent are conveyed through human movement. We know for example that in SL, the spatio-temporal variability of signs can be used to inflect the nature of a sentence and enhance the global expressiveness and style of the virtual signer. However, small spatial or temporal variations may deeply alter the meaning of a sentence.

No field has studied character movement more intently than the performing arts, since their prime goal is to create visually affective and believable characters capable of communicating meaning and emotion to an audience. Therefore, the theatrical body movements can be of interest and employed as a source of inspiration in the understanding of expressive human movement it-self, and in turn exploited in the synthesis of avatars' expressive movements. The reasons behind this idea are threefold:

(i) In the creation of a theater act it is required to develop a deep understanding of "the language of gesture" [65], since it is through movement/gesture that an actor transforms feelings, emotions, intentions, and passions into performance and meaning. By analyzing and understanding the conventions, ideas and techniques employed by theater actors while creating and embodying a character, we may be able to apply similar principles while designing virtual characters with rich emotional expressions.
(ii) While in stage, every movement is deliberately chosen and executed to induce/involve the audience with emotion [4], and thus make every character in scene to be perceived as believable. By using TG as the knowledge base of a motion synthesis system, it is likely that any virtual character will also be perceived as believable and hence the user will be part of a very engaging and meaningful interactive experience.
(iii) In physical theater, the body and its movement are both the center of attention and the center of the theater making process [66]. As spectators, we invest every performer's action and gesture with significance, meaning and emotional/affective content. By studying and analyzing theatrical gestures, we think it is possible to gain an additional insight on how meaning and emotions are conveyed through movement.

5 Conclusion

We have examined in this article the different challenges posed by the animation of advanced expressive virtual characters, according to different aspects that are essential to enhance the believability, comprehension, and responsiveness properties. While data-driven animation techniques clearly show the best believable results, a lot of improvements are still mandatory to fulfill the requirements of theatrical gestures and sign languages production which both require highly nuanced variations, while keeping a strong semantic. Among others, motion capture and corpus building are difficult issues which require significant studio time with experts, and are very costly in post processing. The design of a new computer specification language, including some reactivity in the specification of gestures, and controllers that respect the adaptive motor program constraints should enable the synthesis of responsive gestures. Incorporating both procedural and data driven models with machine learning capabilities is still very challenging, since it allows to combine the generation of realistic movements while giving the possibility to manipulate parameterized behaviors, thus leading to a better control of the synthesis. Finally, the usability and acceptability of virtual characters to the expert communities should also be evaluated thoroughly, notably through the help of professional actors and native signers. Though those issues have recently attracted the attention of several research groups, a lot remain to be done before comprehensive, believable and reactive avatars can be truly effective in our everyday life virtual environments.

Acknowledgements This work is part of the *Ingredible* project, supported by the French National Research Agency (ANR).

References

1. M. Thiebaux, S. Marsella, A.N. Marshall, M. Kallmann, Smartbody: behavior realization for embodied conversational agents, in *Proceedings of AAMAS'08—Volume 1*, Richland, SC (2008) (pp. 151–158)
2. J. Bates et al., The role of emotion in believable agents. Commun. ACM **37**(7), 122–125 (1994)
3. Y. Jung, A. Kuijper, M. Kipp, J. Miksatko, J. Gratch, D, Thalmann, Believable Virtual Characters in Human-Computer Dialogs. In *EUROGRAPHICS*, Llandudno, United Kingdom (2011)
4. M. Neff, Lessons from the arts: what the performing arts literature can teach us about creating expressive character movement, in *Nonverbal Communication in Virtual Worlds: Understanding and Designing Expressive Characters* (ETC Press, 2014)
5. K. Elam, *The Semiotics of Theatre and Drama* (Routledge, 2002)
6. S.C Marsella, S.M. Carnicke, J. Gratch, A. Okhmatovskaia, A. Rizzo, An exploration of delsartes structural acting system, in *Intelligent Virtual Agents* (Springer, 2006), pp. 80–92
7. S. Gibet, T. Lebourque, P.F. Marteau, High level specification and animation of communicative gestures. J. Visual Lang. Comput. **12**, 657–687 (2001)
8. A. Bogart, T. Landau, *The Viewpoints Book: A Practical Guide To Viewpoints and Composition*. Theatre Communications Group (2005)
9. R.E. Johnson, S.K. Liddell, Toward a phonetic representation of signs: sequentiality and contrast. Sign Lang. Stud. **11**(2), 241–274 (2011)
10. K. Duarte, Motion capture and avatars as portals for analyzing the linguistic structure of signed languages, in *PhD thesis, Université de Bretagne Sud* (2012)
11. A. Kendon, Gesticulation and speech two aspects of the process of utterance, in *The Relation Between Verbal and Nonverbal Communication* (1980), pp. 207–227
12. D. McNeill, *Hand and Mind—What Gestures Reveal about Thought* (The University of Chicago Press, Chicago, IL, 1992)
13. S. Kita, I. van Gijn, H. van der Hulst, Movement phase in signs and co-speech gestures, and their transcriptions by human coders, *Proceedings of the International Gesture Workshop on Gesture and Sign Language in Human-Computer Interaction*, vol. 1371, Lecture Notes in Computer Science (Springer, London, 1997), pp. 23–35
14. C. Awad, N. Courty, K. Duarte, T. Le Naour, S. Gibet, A combined semantic and motion capture database for real-time sign language synthesis, in *IVA* (2009), pp. 432–438
15. R. Laban. *The Mastery of Movement* (Plays Inc. 1971)
16. V. Maletik, *Body, Space, Expression: The Development of Rudolf Laban's Movement and Dance Concepts* (Mouton de Gruyte, New York, 1987)

17. D.M. Chi, M. Costa, L. Zhao, N.I. Badler, The EMOTE model for effort and shape, in *SIGGRAPH* (2000), pp. 173–182
18. L. Zhao, N.I. Badler, Acquiring and validating motion qualities from live limb gestures. Graph. Models **67**(1), 1–16 (2005)
19. L. Torresani, P. Hackney, C. Bregler, Learning motion style synthesis from perceptual observations, in *Advances in Neural Information Processing Systems* (2006), pp. 1393–1400
20. D. Bouchard, N.I. Badler, Semantic segmentation of motion capture using laban movement analysis, in *Intelligent Virtual Agents* (*IVA 2007*) (2007), pp. 37–44
21. M. Karg, A. Samadani, R. Gorbet, K. Kühnlenz, J. Hoey, D. Kulic, Body movements for affective expression: a survey of automatic recognition and generation. T. Affect. Comput. **4** (4), 341–359 (2013)
22. A. Kleinsmith, N. Bianchi-Berthouze, Affective body expression perception and recognition: a survey. T. Affect. Comput. **4**(1), 15–33 (2013)
23. N. Eshkol, J. Harries, Ewmn: Eshkol-wachman movement notation (2011). Accessed 10 Mar 2015 [Online]
24. J.G. Harries, Symmetry in the movements of T'ai Chi Chuan. Comput. Math Appl. **17**(46), 827–835 (1989)
25. N. Eshkol, *The Hand Book: The Detailed Notation of Hand and Finger Movements and Forms*. [Tel Aviv]: Movement Notation Society, 1971
26. O. Teitelbaum, T. Benton, P.K. Shah, A. Prince, J.L. Kelly, P. Teitelbaum, Eshkol-wachman movement notation in diagnosis: the early detection of Asperger's syndrome. Proc. Natl. Acad. Sci. USA **101**(32), 11909–11914 (2004)
27. J. Tanenbaum, M. Seif El-Nasr, M. Nixon, Basics of nonverbal communication in the physical world, in *Nonverbal Communication in Virtual Worlds: understanding and Designing Expressive Characters* (ETC Press, 2014)
28. M. Nixon, P. Pasquier, M.S. El-Nasr, Delsartmap: applying delsartes aesthetic system to virtual agents, in *Intelligent Virtual Agents* (Springer, 2010), pp. 139–145
29. W.C. Stokoe, *Semiotics and Human Sign Language*. Walter de Gruyter Inc. (1972)
30. D. Brentari, *A Prosodic Model of Sign Language Phonology* (MIT Press, Cambridge, MA, 1999)
31. S. Prillwitz, R. Leven, H. Zienert, T. Hanke, J. Henning, *Hamburg Notation System for Sign Languages—An Introductory Guide* (University of Hamburg Press, 1989)
32. T. Johnston, The lexical database of AUSLAN (Australian Sign Language), in *Proceedings of the First Intersign Workshop: Lexical Databases*, Hamburg (1998)
33. K. Duarte, S. Gibet, Heterogeneous data sources for signed language analysis and synthesis: the signcom project, in *Proceedings of LREC'10* (ELRA, 2010)
34. D. Vernon, G. Metta, G. Sandini, A survey of artificial cognitive systems: implications for the autonomous development of mental capabilities in computational agents. Trans. Evol. Comp **11**(2), 151–180 (2007)
35. W. Duch, R.J. Oentaryo, M. Pasquier, Cognitive architectures: where do we go from here? in *Proceedings of the 2008 Conference on Artificial General Intelligence 2008: Proceedings of the First AGI Conference*, Amsterdam, The Netherlands. (IOS Press, The Netherlands, 2008) pp. 122–136
36. F. Devillers, S. Donikian, F. Lamarche, J.F. Taille, A programming environment for behavioural animation. J. Vis. Comput. Animation **13**(5), 263–274 (2002)
37. J. Cassell, J. Sullivan, S. Prevost, E.F. Churchill, *Embodied Conversational Agents* (The MIT Press, 2000)
38. A. Kranstedt, S. Ko, I. Wachsmuth, MURML: a multimodal utterance representation markup language for conversational agents, in *Proceedings of the AAMAS02 Workshop on ECA* (2002)
39. H. Noot, Z. Ruttkay, Variations in gesturing and speech by gestyle. Int. J. Hum.-Comput. Stud. **62**(2), 211–229 (2005)
40. B. Hartmann, M. Mancini, S. Buisine, C. Pelachaud, Implementing expressive gesture synthesis for embodied conversational agents, in *Gesture Workshop* (Springer, 2005)

41. C. Pelachaud, *Studies on Gesture Expressivity for a Virtual Agent*, vol. 63(1) (Springer, 2009)
42. H. Vilhjálmsson, N. Cantelmo, J. Cassell, N. Ech Chafai, M. Kipp, S. Kopp, M. Mancini, S. Marsella, A.N. Marshall, C. Pelachaud, Z. Ruttkay, K.R. Thórisson, H. van Welbergen, R.J. van der Werf, The behavior markup language: recent developments and challenges, in *Intelligent Virtual Agents, 7th International Conference, IVA 2007, Paris, France, September 17–19, 2007,* (2007), pp. 99–111
43. S. Kopp, B. Krenn, S. Marsella, A.N. Marshall, C. Pelachaud, H. Pirker, K.R. Thórisson, H.H. Vilhjálmsson, Towards a common framework for multimodal generation: the behavior markup language, in *Intelligent Virtual Agents, 6th International Conference, IVA 2006, Marina Del Rey, CA, USA, August 21–23* (2006), pp. 205–217
44. D. Tolani, A. Goswami, N.I. Badler, Real-time inverse kinematics techniques for anthropomorphic limbs. Graph. Models **62**(5), 353–388 (2000)
45. S. Kopp, I. Wachsmuth, Synthesizing multimodal utterances for conversational agents. J. Vis. Comput. Animation **15**(1), 39–52 (2004)
46. A. Heloir, M. Kipp, Real-time animation of interactive agents: specification and realization. Appl. Artif. Intell. **24**(6), 510–529 (2010)
47. M. Gillies, X. Pan, M. Slater, Piavca: a framework for heterogeneous interactions with virtual characters, in *Intelligent Virtual Agents, 8th International Conference, IVA 2008, Tokyo, Japan, September 1–3, 2008. Proceedings*, (2008) pp. 494–495
48. L. Kovar, M. Gleicher, F. Pighin, Motion graphs. ACM Trans. Graph. **21**(3), 473–482 (2002)
49. M. Brand, A. Hertzmann, Style machines, in *ACM SIGGRAPH 2000* (2000) pp. 183–192
50. A. Hertzmann, Machine learning for computer graphics: a manifesto and tutorial, in *Computer Graphics and Applications, 2003.* (IEEE, 2003), pp. 22–36
51. K. Grochow, S.L. Martin, A. Hertzmann, Z. Popovisć, Style-based inverse kinematics. ACM Trans. Graph. **23**(3), 522–531 (2004)
52. E. Hsu, K. Pulli, J. Popović, Style translation for human motion, in *ACM Transactions on Graphics (TOG)* (ACM, 2005), pp. 1082–1089. (Volume 24:3)
53. O. Arikan, D.A. Forsyth, J.F. O'Brien, Motion synthesis from annotations. ACM Trans. Graph. **22**(3), 402–408 (2003)
54. M. Müller, B. Andreas, H.P. Seidel, Efficient and robust annotation of motion capture data, in *Proceedings of the ACM SIGGRAPHEurographics Symposium on Computer Animation* (2009), pp. 17–26
55. C. Rose, B. Bodenheimer, M.F. Cohen, Verbs and adverbs: Multidimensional motion interpolation using radial basis functions. IEEE Comput. Graph. Appl. **18**, 32–40 (1998)
56. M. Stone, D. DeCarlo, O. Insuk, C. Rodriguez, A. Stere, A. Lees, C. Bregler, Speaking with hands: creating animated conversational characters from recordings of human performance. ACM Trans. Graph. **23**(3), 506–513 (2004)
57. S. Jörg, J.K. Hodgins, A. Safonova, Data-driven finger motion synthesis for gesturing characters. *ACM Trans. Grap.* **31**(6), 189:1–189:7 (2012)
58. A. Safonova, J.K. Hodgins, Construction and optimal search of interpolated motion graphs, in *ACM SIGGRAPH 2007 Papers*, SIGGRAPH'07 (2007)
59. J. Starck, A. Hilton, Surface capture for performance-based animation. IEEE Comput. Graph. Appl. **27**(3), 21–31 (2007)
60. S. Gibet, N. Courty, K. Duarte, T. Le Naour, The signcom system for data-driven animation of interactive virtual signers: methodology and evaluation. Trans. Interact. Intel. Syst. **1**(1), 6 (2011)
61. A. Héloir, N. Courty, S. Gibet, F. Multon, Temporal alignment of communicative gesture sequences. Comput. Animation Virtual Worlds **17**, 347–357 (2006)
62. A. Héloir, S. Gibet, A qualitative and quantitative characterisation of style in sign language gestures, in *Gesture in Human-Computer Interaction and Simulation, GW 2007,* Lisboa, Portugal. Lecture Notes in Artificial Intelligence (Springer, 2009)
63. Ed Hooks, *Acting for Animators* (Routledge, 2013)

64. L. Bishko, Our emphatic experience of believable characters, in *Nonverbal Communication in Virtual Worlds: Understanding and Designing Expressive Characters* (ETC Press, 2014)
65. J. Lecoq, J.G. Carasso, J.C. Lallias, D. Bradby, *The Moving Body (Le Corps Poétique): Teaching Creative Theatre*. Methuen Drama Modern Plays. Bloomsbury Academic (2009)
66. S. Murray, J. Keefe, *Physical Theatres: A Critical Introduction* (Taylor & Francis, 2007)

Task Modelling for Reconstruction and Analysis of Folk Dances

Katsushi Ikeuchi, Yoshihiro Sato, Shin'ichro Nakaoka, Shunsuke Kudoh, Takahiro Okamoto and Hauchin Hu

Abstract Intangible cultural assets such as folk dances and native languages have been disappearing day-by-day. It is important to develop new methods to preserve such assets. Toward this goal, this paper focuses on how to preserve folk dances as performances by humanoid robots. This new method provides not only preservation of such folk dances, but also understanding of dance structures. We demonstrate this preservation method on a humanoid by using Japanese and Taiwanese folk dances. We also explain how such demonstrations provide new insights to folk dance, which leads to interdisciplinary studies of Taiwanese folk dances.

1 Introduction

Cultural assets have been disappearing day-by-day. Tangible assets, such as temples and statues, have been destroyed due to natural disasters such as Tsunami and earthquake or man-made disasters such as the Taliban's destruction of the Bamiyan

Y. Sato · T. Okamoto · H. Hu
The University of Tokyo, Tokyo, Japan
e-mail: yoshi@cvl.iis.u-tokyo.ac.jp

T. Okamoto
e-mail: okagen@cvl.iis.u-tokyo.ac.jp

H. Hu
e-mail: kaikin@cvl.iis.u-tokyo.ac.jp

S. Kudoh
The University of Electro-Communications, Chofu, Japan
e-mail: kudoh@is.uec.ac.jp

S. Nakaoka
National Institute of Advanced Industrial Science and Technology, Tsukuba, Japan
e-mail: s.nakaoka@aist.go.jp

K. Ikeuchi (✉)
Microsoft Research Asia, Beijing, China
e-mail: katsuike@microsoft.com

J.-P. Laumond and N. Abe (eds.), *Dance Notations and Robot Motion*,
Springer Tracts in Advanced Robotics 111,
DOI 10.1007/978-3-319-25739-6_9

Buddhas or ISIS's destruction of various Syrian holy sites. Intangible assets, such as language and folk dances, have the same stories due to cultural mixtures and modernizations. It is important and urgent to develop methods to preserve cultural assets in digital forms.

We have been conducting a project, which we refer to as e-Heritage, to preserve such heritage in digital forms. As for tangible assets, we have developed a pipeline to model various temples and sculptures for displaying them through cloud computers. Examples include Nara big Buddha in Japan and Bayon temple in Cambodia as well as painted tumuli in northern Kyushu Island in Japan [1, 2]. We have also proposed to use such e-Heritage data for archaeological research, which we refer to as Cyber archaeology.

This paper turns the attention to preservation of intangible assets. In particular, we propose to preserve folk dances in the form of performance by humanoid robots. Just for preservation purpose, it may appear to be enough to record those folk dances in videotaping. We argue that such videotaping only provides appearances of folk dances and does not provide the insight into dances, such as which part is important to perform the dances. This understanding is the essence required to inherit the dance. It is also true that such demonstration effort generates symbolic representations of folk dances, which lead in humanity analysis of folk dances.

In the robotics field, many researchers have developed methods to adapt human motion to a humanoid robot. Riley et al. [3] produced a dancing motion of a humanoid robot by converting human motion data obtained by a motion capture into joint trajectories of the robot. For the same purpose, Pollard et al. [4] proposed a method for constraining given joint trajectories within mechanical limitations of the joints. For biped humanoid robots, Tamiya et al. [5] proposed a method that enables a robot to follow given motion trajectories while keeping body balance. Kagami et al. [6] extended the method so that it allows the changes of supporting legs. Yamane and Nakamura [7] proposed a dynamics filter, which converts a physically inconsistent motion into a consistent one for a given body. These works are mainly concern with how to create a new trajectory of a joint within a given physical constraint; there is no attempt to describe global dance structures.

With regard to dance performance, Kuroki et al. [8] enabled an actual biped humanoid to stably perform dance motions that include dynamic-style steps. Nakaoka et al. [9] also developed a similar dancing robot based on the software Choreonoid. These robots are manually coded and no analysis exists. Kawato's group proposes a humanoid robot to learn Okinawa-teodori based on neural network approach [10]. The result is interesting, however due to the bottom-up nature of the learning mechanism, it is difficult to conduct the analysis of dance structure for further preservation purpose. Kosuge et al. [11] proposes a dance-partner robot for western dance. The robot performs excellent dance based on the partner's motion. Okuno's [12] group developed a humanoid robot to step along with the music beat. The motion is limited on stepping actions.

In contrast to these earlier attempts, we focus on describing dances in symbolic representations, i.e. tasks and skills, under the learning-from-observation (LOF) paradigm [13–16]. Tasks and skills are defined to describe what-to-do and how-to-do, respectively. By using tasks and skills, we can separate dance performances into essential and common-to-all-the-performers components, i.e. tasks, and modifiable and personally depending components, i.e. skills. For mapping those tasks and skills to one humanoid platform, the LOF mimics tasks in the exact manner, while it generates modified skills under physical constraints given by the humanoid platform. It is also true that we can utilize such obtained tasks-and-skills on different humanoid platforms relatively easily. Further, by analyzing those tasks we can obtain the global dance structures and deep understanding of the dance.

Section 2 explains our basic LOF paradigm. We will introduce the concept of tasks and skills. A task represents what-to-do and a skill represents how-to-do the task. Section 3 describes how to apply the LOF to making a humanoid robot dance one of the Japanese folk dances, Aizubanda-san dance. We define “key poses” as important postures among dance sequences. A task and a skill can be understood as a transition among two adjacent key poses and its trajectory, respectively. Section 4 describes a method of how to vary robot speed along the variation of music speed. One of the important components in performing dance is its dynamism and its ability to change along with the speed variation of the music. This section explores the method to vary robot’s performing speed under the physical constraints such as the motor capacity. Section 5 considers Taiwanese folk dances. Although Taiwanese indigenous people maintain important positions in pacific indigenous culture, their intangible heritage has been disappearing due to a lack of a writing system. This section provides the symbolic representation of their folk dance for preservation purpose. This section also provides interdisciplinary studies of tribes of indigenous people based on the folk dance representation obtained. Section 6 concludes this chapter.

2 Learning from Observation Paradigm

This section overviews the concept of the learning-from-observation paradigm, which observes human performance, recognizes those motions based on task and skill models, and maps those tasks and skills onto humanoid-robot motions [13]. First we will describe the main concepts, tasks and skills, and then describe how to use these two concepts for mapping.

2.1 Task, Task Model and Task Recognition

We can define a set of states at each domain of human activities. Let’s consider assembly operations of polyhedra. An assembly operation is characterized such as one to create a transition of contact states among two polyhedra. One key question is how to define contact states among two polyhedra. According to Kuhn-Tucker theory, we can classify all possible contact states among two polyhedrons into ten classes (see Fig. 1) [13]. Any contact state can fall into one of the ten possible classes.

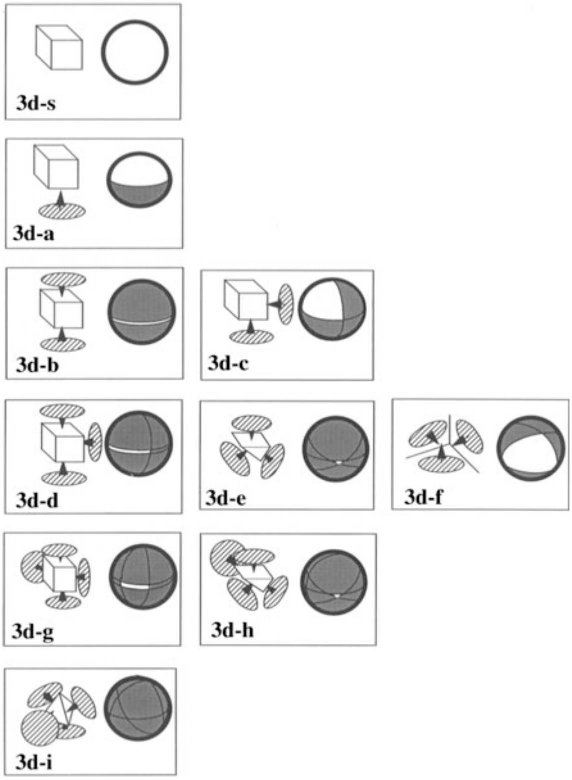


Fig. 1 Kuhn-Tucker's ten contact states. Each state is defined as characteristics of the solution area of simultaneous inequalities, representing movable directions of one polyhedron, the target polyhedron, under the constraint given by the other polyhedron, the constraining polyhedron. For example, in "3d-s" state, the target polyhedron can move all the directions, because the polyhedron is floating in the air and has no constraint from the constraining polyhedron. In "3d-a" state, the target polyhedron sits on the constraining polyhedron. The target polyhedron can only move upward directions, indicated as a *white* northern hemisphere indicating movable directions

We have found that we can define similar state transitions in such action domains as mechanical parts assembly, knot tying [14], and daily human actions [15], as well as human dancing [16]. Here we will explain human dancing domain in details in the later part of this chapter.

A task is defined as an operation to create a state transition from one state to another state. Let's use the same polyhedral world as an example. How many classes of tasks, which we refer to as task models, exist in the polyhedral world. For naïve observation, since we have ten starting states and ten ending states, we would have one hundred state transitions and one hundred task models are necessary. But a careful analysis reveals only thirteen transitions are possible and we only need to prepare thirteen task models. Each task model provides a template of what-to-do, which defines the operation to generate a state transition from one state to another state.

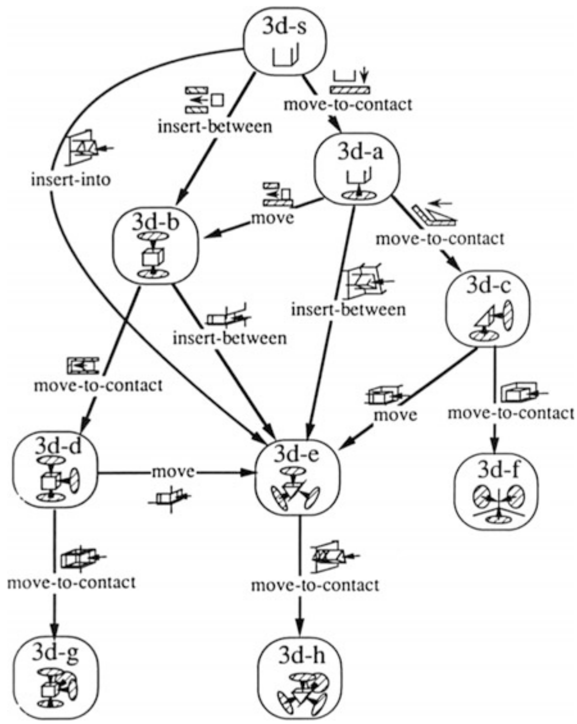


Fig. 2 Thirteen task models in the polyhedral world. Each *arc* represents a state transition. This diagram contains thirteen arcs. To each arc, one task model is defined, which generates corresponding state transition. For example, "move-to-contact" task is assigned to the arc connecting the state transition from "3d-s" state to "3d-a" state

Figure 2 shows the thirteen task models, defined in the polyhedral world. In the figure, each arc represents one state transition. For example, the left top arc represents the transition from the "3d-s" state to "3d-a" state, which is associated with the task model, "move-to-contact."

Task recognition is also used to recognize human actions. From observing continuous human actions, at each moment, a system examines the current state. When a system detects a new state to occur, it analyzes what kind of state transition happens, and retrieves the corresponding arc to retrieve the task model. Namely, the system can infer the necessary action to cause such state transition based on the task model. Simultaneously, the system can segment a continuous motion into an interval corresponding to one task.

2.2 Skill, Skill Model and Skill Recognition

A skill is defined to describe how-to-do the task. One task model only describes what-to-do the task, such as "move-to-contact", and does not describe the details such as where-to-grasp for the move or what-trajectory-to-follow during the move. A skill defines "how-to-do" the task. Each skill is different from each other,

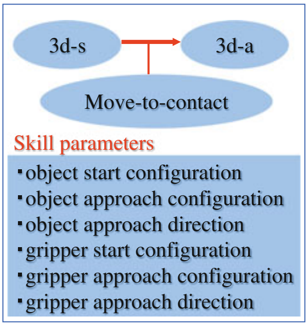


Fig. 3 A task and skill model for "3d-s to 3d-a" task model

depending on that particular task to be done; we have to define a different set of skill parameters, and collectively name them as a skill for the task.

Figure 3 shows an example of a skill model, a set of skill parameters, corresponding to "3d-s-to-3d-a" task model. Here, various detailed configurations of the gripper are defined. The trajectory is also defined as approach directions.

Skill recognition occurs after task recognition finishes. We can pre-determine where to look to obtain such skill parameters. This information is embodies as a demon in each slot of skill parameters. Each demon recovers corresponding parameters from the interval corresponding to the task.

The definition of a skill defined above is one particular example of how-to-do, observed at that particular instance. The how-to-do performance of a dance may vary along trials even by the same performer depending on his/her internal status, such as exciting conditions or exhausted conditions as well as external status such as music tempo. This how-to-do also varies depending on dancers. But, for the time being, for the sake of clarity, we only handle one particular set of skill parameters given by one performer at one particular occasion.

2.3 Task and Skill Mapping

Each robot has different dimensions and physical constraints. We prepare a skill execution module, which records a prototypical trajectory among other things, for each task, given a robot platform. When conducting actions, the system modifies trajectories based on the observed skill parameters, which characterizes the trajectory. Then, each trajectory is connected by considering dynamic balance and collisions.

3 Task Modelling for Dance Reconstruction

This section introduces our earlier work on dancing Aizubanda-san by a humanoid robot by Nakaoka et al. [16]. Here we emphasize the role of task and skill models, given by one performer at one particular occasion. Each part of the human and robot body plays a unique role for dancing. The role of the lower body is to stably support the upper body, while the role of the upper body is to represent various dance configurations. The role of the middle body, the waist, is to adjust motions between the upper and the lower body. Thus, we define different class of task and skill models for each part.

3.1 Task and Skill Models for Lower-Body

3.1.1 Modelling

The main purpose of the lower body is to stably support the whole body. It is also true that the lower body is hidden by Kimono in Japanese fold dance. For these two reasons, we limit the freedom of the lower body, and only allow three states: left foot standing, right foot standing, and both feet standing. Theoretically, three state transitions, and thus, three task models are possible. However, due to the hardware limitation of our humanoid robot, we consider only two state transitions and, as the results, two task models, as shown in Fig. 4. We do not allow the direct transit task between left and right standing states.

We prepare task recognition system for dance motion based on two task models. From the continuous motion, at each moment, the system monitors the distance between the foot and floor and decides the current state among the three states. When it detects a new state, it segments the interval and recognizes which state transition occurs; the system recognizes occurrence of a task, which is defined as a state transition. In particular, in Japanese folk dance, the beginning of one task of the lower body is always one both-feet-standing state. Then, either left or right foot standing state occurs. Finally, the both-feet-standing state occurs at the end. The interval between the two both-feet-standing states corresponding one task. Here, the interval corresponding to the task includes the starting both-feet-standing, but does not includes the ending both-feet-standing.

A skill, a set of skill parameters, defined to each task, is obtained from a task interval. The system cuts out an observation sequence into intervals corresponding to tasks. From analyzing one interval, corresponding to one task, the system obtains

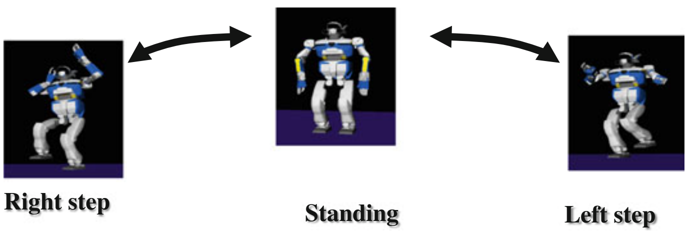


Fig. 4 Three states and two task models for the lower body

appropriate skill parameters to the task. In particular, for stability of robot bodies, we prepare a prototypical trajectory of the foot for each task. This trajectory is characterized using two skill parameters, highest point of a trajectory and the width of a foot standing. Another skill parameter is the duration of a task. Those skill parameters are obtained from the task interval.

3.1.2 Mapping

First, a sequence of tasks is generated from the observation. Then, at each task, from the observed parameters, the prototypical trajectory is modified based on the observed skill parameters. Once the foot trajectory is obtained, we can determine the motion of the lower body of a humanoid robot by solving the inverse kinematics of lower body of a humanoid robot. Here, we considers a method of self-collisions by using an iterative modification method. As for dynamic balancing, we rely on AIST real time balancer for execution. This module also modifies next foot-step on-line. The details of the modification method can be found in [16].

3.2 Task and Skill Models for Upper-Body

3.2.1 Modelling

In order to get a hint regarding upper body state, we have asked one of the dance teachers to draw a sketch of the upper body motion. She provided us the drawings shown in Fig. 5. Apparently, this sketch provides a rough idea of how to perform this folk dance.

The next question was how to extract those states. Originally, we assumed that those states correspond to points of brief stops along dance performance. After extracting brief stops along dance motions, we realized that too many brief stops were extracted. Among those brief stops, some were important points that correspond to a new state, while others were accidental ones.

Fig. 5 Drawings given by a dance teacher

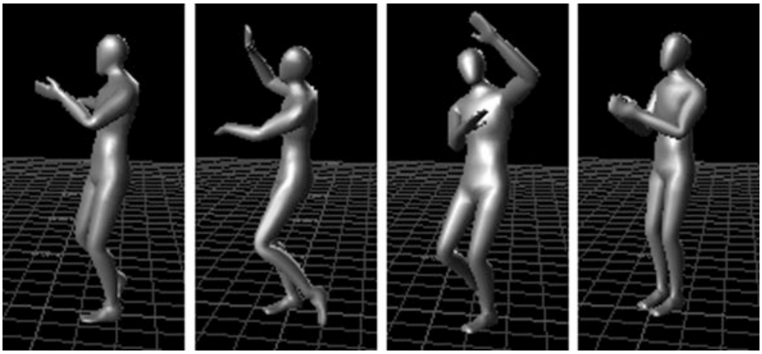
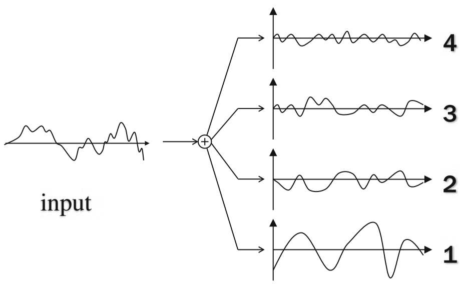


Fig. 6 Postures automatically extracted

We introduce music beats for discriminating those important stops from the accidental ones. If a brief stop occurs at the beat point, we consider the stop as an important and intentional one. Otherwise, we consider the stop as an accidental non-intentional one. The result is shown in Fig. 6. Figures 5 and 6 corresponds to each other. This result supports our assumption that a brief stop on a music beat is an important intentional one. We name the posture corresponding to an important stop as a key pose. We can define upper body tasks as transition between two key poses. Again, one task is defined as the interval, which starts from one key pose and ends immediately before the next key pose.

A skill in the upper body motion consists of trajectories of each joints and the duration of a task. In order to represent an appropriate trajectory, we employ a hierarchical B-spline curve. See Fig. 7. First, a trajectory is represented as a B-spline curve at a certain sampling interval. The residue between the original curve and the B-spline curve is calculated and the residue is represented as finer B-spline curves at a double-sampling interval. This process is iteratively repeated and ends up a hierarchical B-spline representation [17].

Fig. 7 Hierarchical B-spline curve



3.2.2 Mapping

Mapping tasks to a humanoid is relatively straightforward. In any case, a pair of postures corresponding to a task should be exactly mimicked. This is because such postures characterize the dance.

Skill mapping is more complicated. Due to motor limitation, sometimes it is impossible to mimic the exact trajectory given by a human performer. We have to modify the trajectory in order for a robot to be able to perform it. During the execution of a task, we try to mimic the trajectory as much as possible within the limitation of motor capacities. If we detect motor capacity exceeds its limit at a certain point along the trajectory, we remove the highest degree of hierarchical B-spline component, generate a new trajectory, and check whether the new trajectory is within the motor limitation. If the new trajectory is not within motor limits, we repeat this process iteratively until all the points along a new trajectory are within the motor capability.

3.3 Synchronization of Lower and Upper Body Motions

The upper body motion and lower body motion are independently generated. The synchronization is done at each key pose. In other word, one task of upper body corresponds to one task of lower body. By using this schema, synchronization is conducted between lower body and upper body. The next issue is stability. Originally, the lower body motion was generated so as to be stable. However, by adding upper body motion, dynamic stability is lost. We will adjust the position of the wrist so that the resulting ZMP is inside of a foot print. This is handled by AIST dynamic simulator. Figure 8 shows the result of a dancing sequence generated by the humanoid robot.

Fig. 8 Humanoid dance

4 Skill Modelling for Catching up Music Speed

The method described in the previous section provides a dancing humanoid based on observed speed of the dance. Skills with trajectories of joints, are specific to one particular performance by one dancer at one occasion. Dance performance, however, depends on the speed of the music at each different occasion. This is an interesting and exciting feature of live performance of dance and music; one day, the music and dance speed are higher than usual; the other day, the music and dance speed are slower than usual. It is desirable for a robot to modify the performance speed along with the accompanying music speed. In this section, we will explore the method to modify trajectories depending on the music speed.

When the music speed is slower than the predetermined music speed, the robot simply slows down its motions; there is no difficulty in this case. On the other hand, if the music speed is faster than the original one, there may be a difficult point along the performance to follow the music due to motor limitations. The robot may have to omit details in trajectories to catch up with the music speed. This issue also occurs in a human performer because we, human dancers, also have similar limitations of performance coming from the limitation of muscle power. It is desirable for a robot to omit details in a similar way as a human dancer does. Thus, this section observes how a human dancer solves the limitation issue and applies the same strategy to the robot [18].

4.1 Lower-Body Skills

In order to observe human performance, we ask a couple of dancers to perform the same dance at different music speeds, and observe which components are maintained, and which components vary. In other word, we observe how each skill varies depending on the music speed. We focus four components: timing of steps, widths of steps, maximum speed of steps, and trajectories of steps.

Figure 9 shows the timing of variation of step timings. Apparently, steps corresponding to key poses have lower variations and steps far from key poses have high variations. Namely, the key pose steps are conducted at particular timings independent of music speeds so that key pose steps occur at the predetermined music beats. Un-related steps are conducted arbitrarily probably because of energy saving purpose.

Figure 10 shows the widths of steps. The step length is maintained independent of music speed. This fact can be understood as dances are performed in predetermined areas, say from the left side of a stage to the center of the stage; each step should have the same length independent of the music speed.

Figure 11 shows the maximum speed of a foot. Again, the maximum speed is maintained. Namely, a dancer always conducts dances using full energy to show quick actions. Of course, at the extreme case around two times faster, we can

Fig. 9 Step timing

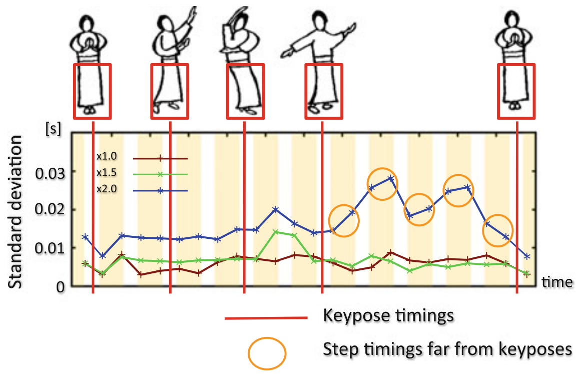


Fig. 10 Widths

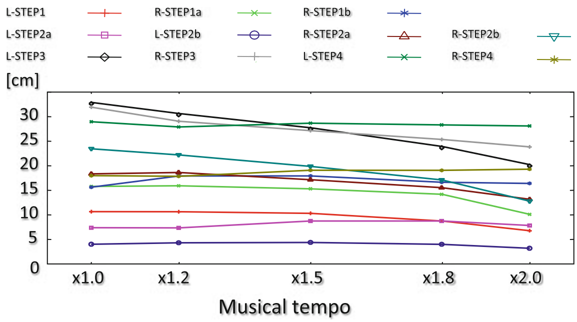
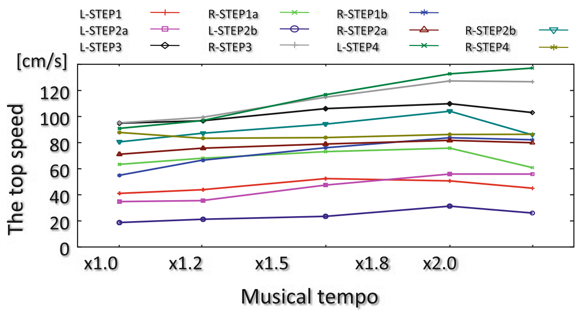


Fig. 11 Maximum speed



observe the break down probably due to the limitation of muscle output. But up to that point, all the dancers try their best to maintain the maximum speed.

Then, which components are used to remove the details for catching up? In fact, the trajectory is the component used for this purpose. Figure 12 shows how foot trajectories vary depending of the music speed. When the music speed increases, the trajectory becomes compact.

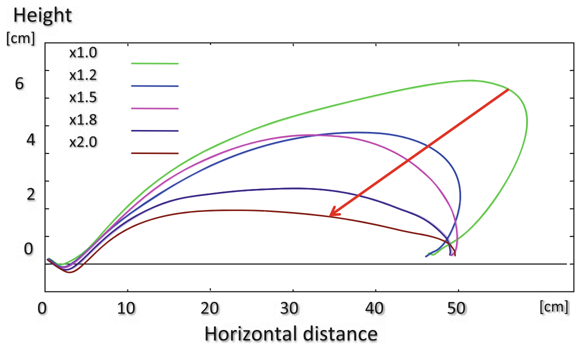


Fig. 12 Foot trajectories

From this observation, we can obtain the following generation strategy of lower body motion:

- The timings of steps around key poses should be maintained.
- We maintain the width and maximum speed of each step.
- Steps far from the key poses can be generated at arbitrary timings.
- The trajectories become compact along with the music speed. For robot implementation, depending on the music speed, we reduce the highest point in the trajectories for the lower body motions.

4.2 Upper-Body Skills

The key pose timing of the upper body motion is also maintained in varying music speed. In the same way as the lower body, the trajectory of the upper body varies depending on the music speed. The trajectory becomes compact when the music speed increases.

In order to represent this compactness, we utilize the hierarchical B-spline in the same way as in the previous section. At each necessary speed, we generate a trajectory and see whether the trajectory is within the limit of the motor. If some part exceeds the limit, we gradually remove higher order representations.

We implemented this varying speed of dance performance using Donpan dance, one of Japanese folk dances in Tohoku area. Unfortunately, we cannot effectively show the result in a figure. If interested, please take a look at our web site for demonstrations. In the video, a robot dances the Donpan at 0.8 speed, standard speed, and 1.2 speed.

5 Task Models, Labanotation and Taiwanese Dances

The previous section explains how to make a humanoid robot to perform folk dances based on the task analysis. This section provides a new method to use such task models for analyzing folk dances [19]. We analyze Taiwanese indigenous folk dances and make comparisons among them.

Taiwanese indigenous people are considered the origin of Austronesians who began to spread out among the Pacific islands around BC3000. They were the first ones to settle down on the Taiwanese island. Currently, 14 tribes are recognized: Amis, Kavalan, Sakizaya, Puyuma, Paiwan, Rukai, Bunun, Atayal, Seediq, Taroko, Sisiyat, Tsou, Thao, and Tao. They live in the central, eastern, and southern mountain areas of Taiwan Island. They have a unique culture from long traditions. Unfortunately however, due to lack of a writing system, their culture has been lost day-by-day. This is the motivation for us to choose their folk dance for robotic preservation.

5.1 Representing Festival Dances Using Labanotation

We collected the videos of all the festival dances performed by them. Each tribe has a wide variety of dances. We can classify them into three classes: festival dance, life dance and work dance. The festival dances are performed for praying to ancestors, and for sunshine or rain. The life dances are performed at weddings and welcoming or funeral ceremonies. The work dances are performed for the occasions of hunting, fishing and farming; praying for good harvests. Among these three classes, we chose to record festival dance. All the tribes have at least one festival dance. Thus, we can use such festival dance for comparing their culture. It is also true that all the festival dances are attended by all the villagers and performed as a group dance for representing the unity of the tribe. We can consider that such festival dance maintain an important position in their culture.

The next issue is how to convert such dances into symbolic representations. In our various sections we use dance teachers' drawings as the starting point of detecting key poses. We argue that the posture of the key pose is the central representation of the dance. In Japanese folk dance, we can ask dance teachers to perform dances in motion capture system; we can create such notation automatically. However, for Taiwanese dance, we only have videos, we have to manually convert such videos into Labanotations by human detection of these key poses. Here, all the dancers shake hands; thus, the freedom of the upper body is rather limited. We have concentrated on the lower body motion.

We have converted all the videos into labanotations. Figure 13 summarize all 4 step dances. Here, CW means that the group dance is performed while rotating in the clockwise direction, and CCW means the counter-clockwise direction; Fig. 14 shows the labanotation of all 2 step dances.

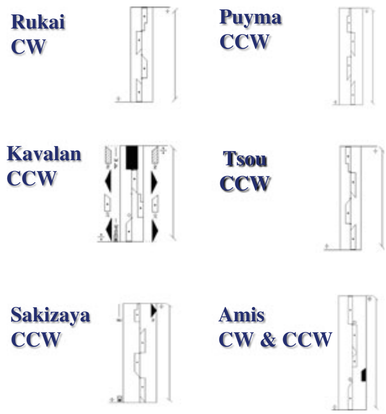


Fig. 13 Labanotation of 4 step dances

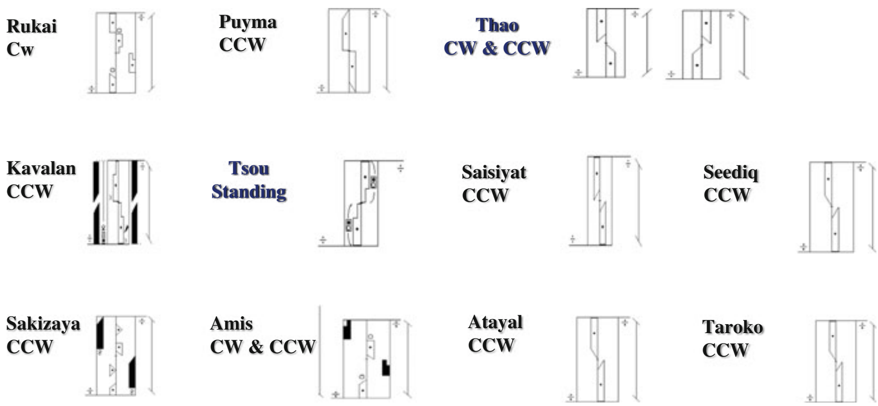


Fig. 14 Labanotation of 2 step dances

Figure 15 shows the classification of those dances. Some groups such as Paiwan and Rukai, have both 4 step dances and 2 step dances. Other groups such as Thao and Saisiyat, have only 2 step dances. Bunun and Tao do not have step dances. As for dancing directions, some groups such as Paiwan and Rukai rotate clockwise, while other groups such as Kavalan and Sakizaya rotate counter-clockwise. Amis rotates in both directions.

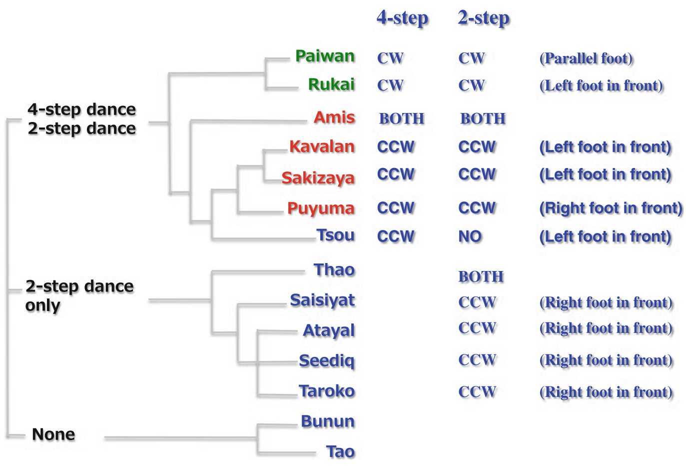


Fig. 15 Dance classification

5.2 Reconstruction Based on Labanotation

Labanotation has been used for recording various performances. In theory, a human dancer can perform a dance based on its Labanotation. However, it is not true that such Labanotation contains all the necessary information to reconstruct dance motions by a robot. In order to verify this point, we try to reconstruct robot motion based on Labanotation.

Through this effort, we realize that Labanotation records only static states of each body part, such as the final foot point or the angle of an arm at that state. In our terminology, this corresponds to a state or a key pose. Labanotation score denotes transitions of states, thus, Labanotation denotes a chain of tasks, what-to-do sequence.

As seen in the previous section, a skill for each task is also necessary for completing the robot motion corresponding to the task. In particular, for a dance, trajectory for a task is one of the most important components. For a human dancer, trajectory is inserted without any intention. However, for a robot it is necessary to insert the trajectory. When we use a robot simulator, Choreonoid, it automatically inserts such a trajectory. We input those tasks given by the previous section and skills, in-between trajectories, are automatically generated by Choreonoid simulator. Figure shows the example of Paiwan dance generated by Choreonoid simulator.

In this section, this trajectory is generated automatically based on a prototype. Since festival dances are played as a group, this standard method works. However, other type of dance, such as hunting dance need to represent subtle differences of dance. Along a trajectory, variance of acceleration should be represented to denote smooth or sudden motions. Laban effort represents such dynamic information. Further investigation to record such dynamic information using Laban effort, and to map robot motion is necessary.

5.3 Dance Analysis Based on Labanotation

We can make a dendrogram of indigenous tribes based on dance features. In this chapter we will examine how this dendrogram relates to other dendrograms based on other features. We will examine DNA analysis, language analysis, and social institutions.

5.3.1 Comparison with DNA Analysis

There are many DNA analyses. Here, we decide to use mtDNA, mitochondrial analysis. In this area, Tajima published the Tribe dendrogram based on his mtDNA analysis as shown in the left side of Fig. 16 [20], while the right side of the figure is our dendrogram based on dance features. The results do not correspond. Another mtDNA dendrogram based on Trejaut's analysis [21] also does not correspond. From this, we can conclude that dance classification has no similarity with tribe classification based on Tajima's and Trejaut's mtDNA analyses.

5.3.2 Comparison with Language

Lee classifies those spoken language into four dialects: Eastern, Southern, Northern and central dialect as shown in the left side of the Fig. 17 [22], while the right side is our dendrogram. From this comparison, we can summarize:

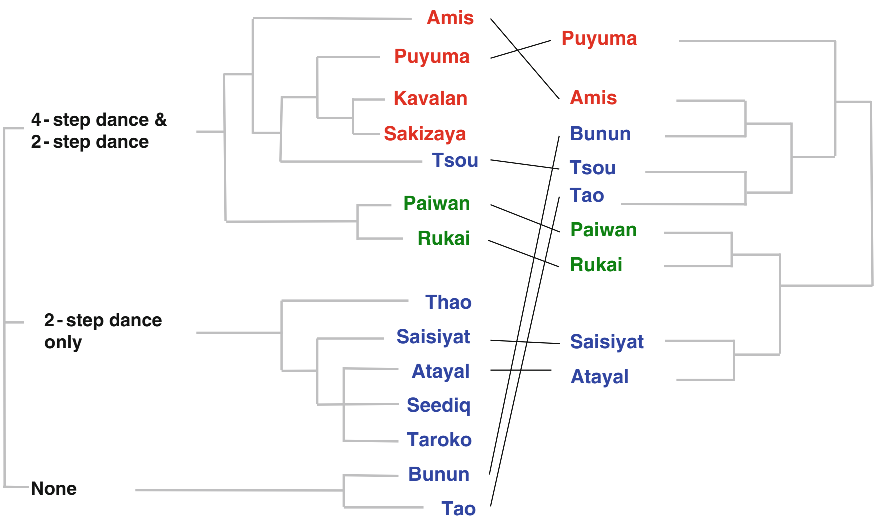


Fig. 16 DNA comparison

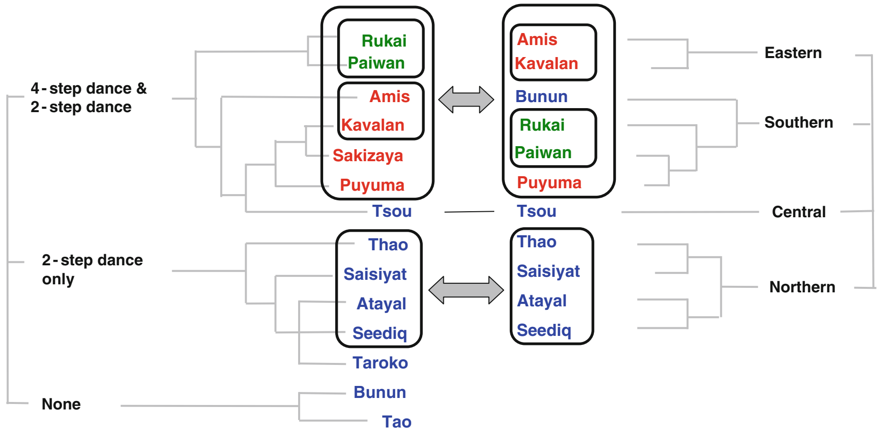


Fig. 17 Language comparison

- Eastern and Southern dialect tribes correspond to 4&2 step tribes.
- Central (Tsou) dialect tribe has 4&2 step dances.
- Northern dialect tribes have only 2 step dances.
- Inner structures of the classification are different from each other.

5.3.3 Comparison with Social Institutions

Their societies can be classified with respect to inheritance in social position and assets. In the aristocratic society, the father's position such as leader, nobility, warrior, and civilian, is inherited by the first son. Paiwan and Rukai belong to the aristocratic society. Other tribes belong to the non-aristocratic society. Namely, anyone who has an ability becomes the leader of the tribe.

Non-aristocratic society can be further classified based on inheritance of assets. In the matrilineal society, such as Amis and Kavalan, the eldest daughter inherits her mother's assets. On the other hand, in the patrilineal society, such as Tso and Thao, the eldest son inherits his father's assets. Usually, men's positions are higher than women's in this society. Based on this Aristocratic/non-aristocratic and matrilineality/patrilineality, we can obtain the tribe dendrogram as shown in Fig. 18.

We can also compare this with our festival dance dendrogram. From this comparison, we can summarize that classification of tribes based on festival dances has a very similar structure with the classification based on social institutions.

As a general summary,

- festival dances have strong relations with social institution
- festival dances also have similar structure with language

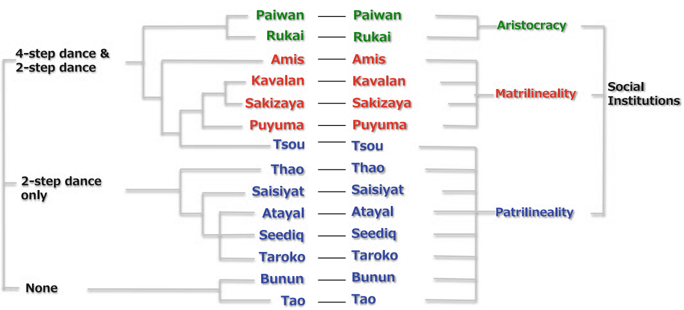


Fig. 18 Dance and social institutions

Since language and social institution of a tribe forms the culture of the tribe, we may be able to say that folk dance also plays an important role in their culture. We can also conclude that culture does not come from human DNA. Culture is formed with social institution, spoken language, and folk dance.

6 Summary

This chapter focuses on the utility of task models for preservation and analysis of folk dances. The main motivation is to preserve intangible heritage in robot demonstration through the paradigm of learning-from-observation with task and skill modeling. First we explore how to define the essential postures in dances, which we referred to as key poses. Transitions between two adjacent key poses are defined as tasks. By using those task models, we were able to reconstruct performances by humanoid robots. We have also classified Taiwanese tribes based on task classifications derived from the Labanotations and pointed out that the dendrogram given by classification of task models has similar structure to the one based on social institution and spoken language.

The remaining issue is the analysis of skills. In particular, dance skills can be described by using spatiotemporal components. Spatio components consist of trajectories of each joint in the space, while temporal components consist of speed along those trajectories. Those skills are also classified into two parts: common components among dancers and specific components to each dancer. With respect to the perspective of skill analysis, Nakaoka's work of Aizu-bandai-san dance, covered in Sect. 4, models one person's one performance both in the spatio-temporal space.

Okamoto's two works mainly consider spatio components in skill modeling. His first work, briefly explained in Sect. 5, considers the case that the same dancer varies his/her performance depending on outer constraints, i.e. music speed [18]. He extracted strategies how to modify skills, which we can consider a meta-skill models. His conclusion is that depending of music speed, each dancer varies the spatio-component, and trajectories of joints become compact. The remaining issue is how to vary such performance depending of his/her internal state, i.e. exciting or exhausted. It is expected that some temporal component such as accretion along trajectories vary.

Analysis on inter performer difference in skills is another important direction. Okamoto's second work considers how much variations occur depending on individual performers [23]. He picked up a ring-throwing as an example. Here, the player has to satisfy the goal that a ring is thrown in a peg. Under this physical condition, he again observes the variation of spatio components. He extracted a few important parameters to characterize trajectories, and then represented parameter distributions as Gaussian distributions. Such inter performance skills should be explore in future.

Acknowledgements This work is, in part, supported by Japan Society for the Promotion of Science, Grants-in-Aid for Scientific Research under Science Research A 23240026.

References

1. K. Ikeuchi, D. Miyazaki, *Digitally Archiving Cultural Objects* (Springer, New York, 2008)
2. K. Ikeuchi, *Digital Restoration of Kyuushu Decorated Tumuli* (Tokyo Shoseki, Tokyo, Japan, 2015). (in Japanese)
3. M. Riley, A. Ude, C.G. Atkeson, Methods for motion generation and interaction with a humanoid robot. in *AAAI and CMU workshop on Interactive Robotics and Entertainmen*t (2000)
4. N.S. Pollard, J.K. Hodgins, M.J. Riley, C.G. Atkinson, Adapting human motion for the control of a humanoid robot, in *ICRA02* (2002)
5. Y. Tamiya, M. Inaba, H. Inoue, Realtime balance compensation for dynamic motion of full-body humanoid standing on one leg. J. Robot. Soc. Japan **17**(2):268–274 (1999). (Japanese)
6. S. Kagami, F. Kanehiro, Y. Tamiya, M. Inaba, H. Inoue, Auto-balancer: an online dynamic balance compensation scheme for humanoid robots, in *The Fourth International Workshop on Algorithmic Foundation of Robotics* (2000)
7. K. Yamane, Y. Nakamura, Dynamics filter—concept and implementation of on-line motion generator for human figures. IEEE Trans. Robot. Autom. **19**(3), 421–432 (2003)
8. Y. Kuroki, M. Fujita, T. Ishida, K. Nagasaka, J. Yamaguchi, A small biped entertainment robot exploring attractive applications, in *ICRA2003* (2003), pp. 471–476
9. S. Nakaoka, A. Nakazawa, K. Ikeuchi, An efficient method for composing whole body motions of a humanoid robot, in *VSMM2004* (2004), pp. 1142–1152
10. G. Cheng, S.H. Hyon, A. Ude, J. Morimoto, J. Hale, M. Kawato, CB: exploring neuroscience with a humanoid research platform, in *ICRA2008* (2008), pp. 1772–1773
11. K. Kosuge, T. Hayashi, Y. Hirata, R. Tobiyama, Dance partner robot-ms dancer, in *IROS2003*, vol. 4 (2003), pp. 3459–3464
12. K. Yoshii, K. Nakadai, Y. Hasegawa, H. Tsujino, K. Komatani, H.G. Okuno. A biped robot that keeps steps in time with musical beats while listening to music with own ears, in *IROS* (2007), pp. 1743–1750
13. K. Ikeuchi, T. Suehiro, Toward an assembly plan from observation i: task recognition with polyhedral objects. IEEE Trans. Robot. Autom. **10**(3), 368–385 (1994)

14. J. Takamatsu, T. Morita, K. Ogawara, H. Kimura, K. Ikeuchi, Representation for knot-tying tasks. IEEE Trans. Robot. **22**(1), 65–78 (2006)
15. K. Ogawara, J. Takamatsu, H. Kimura, K. Ikeuchi, Extraction of essential interactions through multiple observations of human demonstration. IEEE Trans. Ind. Electron. **50**(4), 667–675 (2003)
16. S. Nakaoka, A. Nakazawa, F. Kanehiro, K. Kaneko, M. Morisawa, H. Hirukawa, K. Ikeuchi, Learning from observation paradigm: leg task models for enabling a biped humanoid robot to imitate human dances. Int. J. Robot. Res. **26**(8), 829–844 (2007)
17. T. Shiratori, A. Nakazawa, K. Ikeuchi, Synthesizing dance performance using musical and motion features, in *ICRA2006* (2006), pp. 3654–3659
18. T. Okamoto, T. Shiratori, S. Kudoh, S. Nakaoka, K. Ikeuchi, Toward a dancing robot with listening capability: keypose-based integration of lower, middle, upper body motions for varying music tempos. IEEE Trans. Robot. **30**(3), 771–778 (2014)
19. H. Hu, R. Tseng, C. Lin, L. Ming, K. Ikeuchi, Analysing Taiwanese indigenous folk dances via labanotation and comparing results with interdisciplinary studies, in *Cultural Heritage Documentation, Preservation, and Protection* (Springer, New York, 2014), pp. 196–206
20. A. Tajima, S.S. Shun, I. Pan, T. Ishida, N. Saitou, S. Horai, Mitochondrial DNA polymorphism in nine aboriginal groups of Taiwan. Hum. Genet. **113**(1), 24–33 (2003)
21. J.A. Trejaut, T. Kivisild, J.H. Loo, C.L. Lee, C.L. He, C.J. Hsu, Z.Y. Li, M. Lin, Traces of archaic mitochondrial lineages persist in Austronesian-speaking Formosan populations. PLoS Biol. **3**(8):247 (2005)
22. P. Li, The internal relationships of Formosan languages (2006), pp. 17–20
23. T. Okamoto, T. Shiratori, M. Glisson, K. Yamane, S. Kudoh, K. Ikeuchi, Extraction of person-specific style based on a task model and imitation by humanoid robot, in *IROS2014* (2014), pp. 1347–1354

Dynamic Coordination Patterns in *Tango Argentino*: A Cross-Fertilization of Subjective Explication Methods and Motion Capture

Michael Kimmel and Emanuel Preuschl

Abstract This contribution surveys strategies for analyzing highly coordinated forms of collaborative dance improvisation, based on *tango argentino*. Specifically, we take interest (a) in micro-coordination at the <1 s timescale in dance elements such as steps or rotations, (b) in meso-scale patterns of 2–8 step "figures", and (c) in general enabling macro-patterns maintained throughout, a kind of "grammar" of tango. Across these timescales, dancers supply individual action-readiness, dynamic stability, proper form and connectivity, while jointly "managing" structures of interpersonal coordination such as enabling configurations. Our study engages qualitative and quantitative methods in a dialogue. Starting with micro-genetic elicitation interviews, dancers reported ideomotor concepts, perceptual triggers, and didactic imagery. Besides general (e.g. postural) habits, task-specific forms or techniques, and attentional foci, this yields insights into the interlocking contributions and information flow between tango leaders and followers within units as small as half-steps. The subjective data was then "front-loaded" into a motion-capture study in which six expert couples, fitted with 2 × 21 light-point reflectors, executed various tango techniques. We developed kinematic indicators for individual and interpersonal coordination (degree of coupling, relative movement onset and action timing, role specifics); we also measured geometries underlying various tango tasks and style variations. The combined data suggests a micro-coordination model where dynamic interdependencies enable precise mutually adaptive action. The criss-crossing signals, e.g. when the increasing lability of the leader's torso triggers the follower's leg extension at the beginning of a forward step, suggests task-, phase- and body-part specific contingencies whereby leaders and followers micro-coordinate actions with respect to one another.

M. Kimmel (✉)
Cognitive Science Platform, University of Vienna, Vienna, Austria
e-mail: michael.kimmel@univie.ac.at

E. Preuschl
Institute of Sport Science, University of Vienna, Vienna, Austria
e-mail: emanuel.preuschl@univie.ac.at

J.-P. Laumond and N. Abe (eds.), *Dance Notations and Robot Motion*,
Springer Tracts in Advanced Robotics 111,
DOI 10.1007/978-3-319-25739-6_10

1 Introduction

In recent times scholars have begun to investigate *participatory sense-making* [8, 13] in numerous interaction domains.[1] Pair dance, especially when improvised, presents this human ability at its most complex. Joint improvisation demands a confluence of sensorimotor, interactional, and creative abilities: no small feat. In systems such as tango argentino perfectly coordinated behavior emerges "inter-enactively" between bodies [45], cf. [43]. Dancers dynamically adapt to [17] and actively *co-regulate* [9] each other in a continuous stream of reciprocal causation. Coordinative patterns are thus negotiated in real-time.

Strikingly, these patterns turn out highly synchronized, complementary, and well-formed. As a keystone in the bigger picture of joint improvisation, a theory of dynamic order is therefore needed. Accordingly, in this paper we wish to illustrate key parameters underlying well-formed tango interaction. We shall point out different timescales and ecological scales of order (individual/dyadic) and introduce metrics useful for capturing in sufficient detail complex interaction or, in the future, perhaps even simulating such skills. Our second aim is to advance general methodology by illustrating the essential complementariness of 1st and 3rd person approaches in modeling dynamic order.

As a necessary fundament for quantitative work we need to understand the tango's subjective logic: Geometries and trajectories dancers highlight to their attention reflect crucial determinants of form around which movement organizes. Equally important are the couple's coordinative aims and the informational micro-signals dancers act on. Rote-learned step sequences being absent, dynamic interdependencies between partners—kinetic and/or conventionally semiotic—organize interpersonal coordination. The couple's observable feats arise from cascades of precisely interwoven individual micro-actions, coordinated with one's partner via sensorimotor interdependencies. Thus, criss-crossing informational cues generate dynamic order.

Our subsequent task is to biomechanically model the spatio-temporal and functional coordination couples achieve (cf. [1]). While interaction studies mostly measure simple coordinative macro-variables like relative phasing we embrace a multi-dimensional analysis of individual body-part coordination and interpersonal coordination as a whole, including how individual micro-coordination figures in the larger whole.

[1]Qualitative studies include mother-child interaction [10, 33, 40, 53], collaborative creativity [35, 36], music ensembles [39], flamenco [25], theater improvisation [26], bodywork [22], and psychotherapy [16]. Dynamic systems researchers and scholars of interpersonal synergies have applied quantitative interpersonal coordination measures (cf. [7, 29, 34]). Examples include rugby, basketball, and futsal [4, 30, 48], horseback riding [23, 31], jazz [38, 49] and workplace teams [41].

1.1 A Sketch of Argentine Tango

Tango argentino has been described as dialog [42] with "kinetic connectivity" [28] and an "invitation-response" structure. A leader and a follower connected through an embrace execute almost always non-identical, yet functionally complementary and synchronized actions. Skills are (partly) role-specific. The leader incorporates real-time offerings of the music, the partner, and the dance-floor and suggests a joint movement. The follower responds to this "invitation" in controlled increments, otherwise remaining neutrally poised and action-ready, while providing the good posture, balance, and grounding that it takes for dynamic stability. Besides some leeway for step timing or ornamentation, followers shape only the "how", but not the "what" of the tango, the improvised choices which are the leader's responsibility.

Tango emphasizes perfectly connected joint action with strong rapport between partners, while transducing the music into expression and blending with the overall dance-floor. Its improvisational appeal lies in responding to *affordances* [15] on the spur of the moment, i.e. the sum total of melodic, rhythmical, configurational, kinetic, postural and other perceptual pointers. Every dance thus emerges step-by-step, without planning ahead more than a second at best (apart from professional stage choreographies). Good leaders chain elements at a rate of four or five per second, including walking, rotations, leg crosses, leg displacements/invasions, leg wraps or swings, off-axis techniques, and ornamentations. Within several dances a couple at a *milonga*—the venue where dancers meet—typically remains on the dance-floor one sees hundreds of combinations drawing from a vast pool of possibilities, a "matrix" tango is set on [5] with thousands of states and linkages between them.

Within any chosen task leaders and followers must micro-coordinate their action flow relative to the partner's with precision. Non-delayed responsiveness without overshooting works by virtue of tight sensorimotor coupling. Through the embrace partners establish an ultra-sensitive resonance loop with bidirectional information flow. This loop is immediately responsive thanks to "mutual dynamical entanglement" of bodies [11]. Dancers literally interpenetrate each other through musculoskeletal inter-body chains. An *extended* control loop accrues, enabling one to feel beyond the upper-body interface, via the hips, into the partner's legs [6].

Tango is also noteworthy for formal precision, exhibiting constraints both in individual and interaction patterns: it is *structured* improvisation. Self-imposed restrictions are the condition of safe, fluid, and unambiguously coordinated joint behavior within a musically and spatially evolving environment. We shall later explore these fundamental body-grammatical constraints and the movement repertoires they accommodate. Within this framework, tango faces the learner with a "multi-skill" ecology: Dancers must master interaction management [19], as well as sensory acuity, dexterity, and musicality skills. Leaders additionally develop improvisation-specific skills—knowing how and where to creatively connect basic elements, whilst exploiting unexpected opportunities and correcting mishaps on-the-fly [20]. Trained dancers integrate these skills in real-time [21], achieving a balance between multiple constraints.

1.2 Micro-genetic Explication Interviews

To reflexivize dancers' praxeological knowledge we applied micro-genetic methods: *Explication Interviews* developed by empirical phenomenologists [32, 46, 47] were tuned to the specificities of embodied interaction. This method engages tango teachers and learners in a dialogue about *thin slices* of interaction (approx. 3–5 secs). Strictly avoiding "why" questions, a micro-timeline is constructed and the phases increasingly thrown into relief concerning the "how". The dialogue with the interviewer—himself a dancer—helps the interviewee arrest attention on and progressively dissect the micro-structure of experience. We collected 25 teacher and learner interviews, 7 think-alouds of (semi-)advanced dancers while sparring, and M. Kimmel kept a micro-analytic diary of his progress as a dancer between his 4th and 8th year.

The interviews focalize (a) ideomotor imagery underlying various movement tasks, proper posture, and interpersonal rapport as well as (b) sensory feedback employed for movement control. Specifically, we investigated didactic imagery expressed, e.g., in metaphors and gestures. This spotlight on action-enabling images was expanded into a comprehensive study of perceptual triggers, through which partners coordinate *within and across* basic tango elements.

Our narrowest analytic timescale of dynamic order concerned minimal *action concepts* [37] for steps, pivots, etc. Although such basic "movemes" span only 500 ms or less, dancers readily address their causal-temporal substructure. This becomes evident by having dancers describe kinetic, tactile, visual, and balance-related feedback information received underway that is used for phase-sensitive execution and fine-tuned timing (*micro-affordances*, [19]). As soon as a selected "moveme" begins, dancers pick up feedback indicating that their action is "on track" or requires dynamic correction [51]. Perceived micro-affordances also provide for exact micro-timing and action increments matching the partner's. Since followers recognize onset triggers for half steps or less, each "moveme" can be micro-coordinated with the leader with phase-sensitivity. Our analysis traced the cascading micro-actions whereby the partners reciprocally trigger each other within the resonance loop. Reconstructing this sort of *informational model* requires matching the leaders' and followers' triggers on a timeline.

The next-higher timescale concerns transitions between "movemes", as leaders select the upcoming element. While novices tend to recall continuations from familiar or prototypical situations and choose from this pool, seasoned leaders learn to *feel* continuation affordances ad hoc. They recognize "doables" (in the sense of open degrees of freedom) from the dancefloor situation even when it is new by using its proprioceptive, pressure, balance, and visual signatures. The geometry of the two chests and hips, diagonal or opposite weighted legs, step position, and the torso's twist conjointly specify options. Seasoned dancers thus master an abstract functional logic of tango [52].

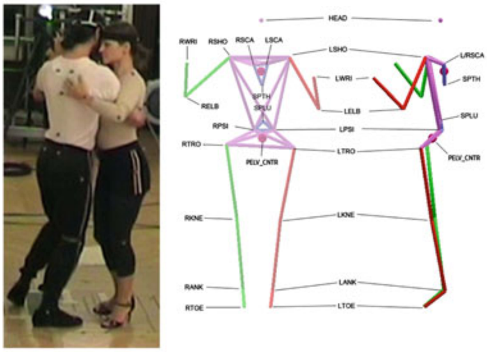
At the timescale of multi-step combinations, we took interest in routines of elements that leaders mentally store as mini-scripts. Frequently, improvisers may utilize these as basic material, e.g. to insert leg wraps, ornamentations, or to truncate and reconnect two scripts. Dancers may also use scripts as learning heuristics to define basic positions and nodes, while progressively modularizing and reshuffling their repertoire [20].

At the highest timescale, we investigated permanent enabling conditions established on entering the dancefloor, which are then actively maintained throughout. The idea is that certain general degrees of freedom must be limited [3] for enabling situated abilities. First, individual dancers internalize patterns like being "in axis" that are conducive to good form (uprightness), action-readiness (*metastable* balance), receptiveness to incoming information and intra-body information flow (inner connection), and muscle efficiency (core activation). Secondly, both partners actively limit degrees of freedom within the ensemble and seek enabling geometries that facilitate interaction management.

1.3 Motion-Capture

Six tango teacher couples, five leading males and six female followers, were invited to participate in a motion-capture study at the Dept. of Sport Sciences of Vienna University. We employed an eight-camera high-speed infrared VICON motion capturing system at 100 frames per second. A set of 21 reflective markers was placed on each dancer, as shown in the stick figure and the corresponding anatomical landmarks in Fig. 1. These markers comprise a tango-specific *body model* that tracks functional features (parts and relations) that teachers judged to "make a difference", but excluding invariants such as head orientation or precise hand positions. To avoid the inevitable occlusions of ventral markers in a close embrace we used lateral and dorsal markers only.

In addition to the 42 point "cloud", including segments and planes between points, *virtual markers* were calculated from selected point arrays: each dancer's pelvis center, approximating the center of gravity (COG) in upright stance, the communication center in the chest (small spheres in the stick figure in Fig. 1), as well as vectors emerging orthogonally from both. Further along, we created summary equations to express interesting dyadic relations like the couple's geometric center or the phase relationship of both dancers' momentary walking directions. Raw data was processed in the application software *Vicon NEXUS*, aided by the modeling software BodyBuilder and its coding language BodyLanguage for biomechanical 3-D models.



HEAD	Head; sagittal suture
L/RSHO	Left / Right shoulder; acromio-clavicular joint
L/RSCA	Left / Right scapula; cross section of the scapula's spine and medial border
SPTH	Thoracic spine; spinous process of the T10
SPLU	Lumbar spine; spinous process of the L4
L/RPSI	Left / Right posterior superior iliac spine
L/RTRO	Left / Right trochanter
L/RKNE	Left /Right knee; lateral epicondylus
L/RANK	Left / Right ankle; lateral malleolus
L/RTOE	Left / Right toe; second metatarsophalangeal joint
L/RELB	Left / Right elbow; lateral epicondyle
L/RWRI	Left / Right wrist, base of the 3rd metacarpal bone

Fig. 1 Lab snapshot, tango body model and marker positions

Two sub-studies were conducted:

- **Study 1: Linear walking in both directions**: The couple walked 4–5 steps six consecutive times for the forward walk and six times for the backward walk. Additionally, a force plate was used to measure the first step's initiation (done once with each partner on the force plate). This study addressed step phases, gait initiation, the two dancers' synchronization over the step cycle, and differences of style.
- **Study 2: (Semi-)free improvisation**: The couples were asked to improvise within eight relatively closely circumscribed tango "themes" for about a minute each, although some offered a certain freedom: 1. walking a square, 2. leg crosses, 3. circling the partner, 4. mutual circling around a joint center, 5. leg flourishes, 6. invasions into the partners steps, 7. off-axis forward leaning, 8. off-axis backward leaning (*cuadrado, cruzado,* two types o*f giro, boleo, sacada, volcada, colgada*). We sampled 40–70 elementary motion units for each theme.

1.4 Making Connections

Nearly all biomechanic studies infuse their inquiry with expertise, yet seldom go about this methodically. Here, micro-genetic *Explication Interviews* afforded full phenomenological systematicity. We then "frontloaded" [14] this subjective data into the biomechanics study. In due course, biomechanic formalization suggested yet deeper subjective inquiry. A genuine dialogue emerged—not least between us, researchers from different backgrounds—with recursive mutual counterchecks and each discipline taking the lead at different times. Kinematic data was systematically arranged to further focalize subjective inquiries; expert phenomenology in return inspired a next round of biomechanic analyses.

Concerning the challenge of finding counterpart connections in two data pools, some are evident, while others require translation of sorts. In the case of floor trajectories, for instance, teachers mention the same rectangle, triangles and circles actually seen (see Sect. 3.1). On the other hand, dancers may rely on, say, pressure related percepts that we cannot directly find in kinematic data, e.g. for determining the mid-step position based on equal pressure on both balls of the feet. In such cases the match had to be inferred from physical and tango-technical principles.

Overall, our endeavor sits well with a surge of multi-method designs combining fine-grained 1st person investigations with neuro-scans and other 3rd person methods (see [12] for an overview).

2 The Dance's Enabling Order

To understand joint improvisation, we must explain how permanent configurative actions enable more task-situated features of the dance system. Successful collaborative improvisation requires, as its enabling condition, general well-formedness, sometimes termed "grammar" by linguistic analogy [2]. I.e., when entering the dance-floor, dancers create and thereupon actively maintain basic structures by monitoring and keeping in their proper range grammar-relevant feedback signatures.

All sophisticated interaction systems require simultaneous levels of parametric organization, any of which missing jeopardizes the overall outcome. Therefore, dancers must reduce degrees of freedom on multiple levels—one angle that explains why tango takes a decade or more to master. In this section we focus on two levels that are most permanent and thus active as a general backdrop: how agents continuously—across tasks—ensure proper individual habits ("body grammar") as well as dyadic habits ("interaction grammar"). Both supply an enabling framework for whatever else happens as strings of tango "vocabulary" unfold.[2]

2.1 Individual Habits

At one level of parametric order, individual dancers contribute under the motto "if I take care of myself, we both benefit": Only with perfect balance, a well-aligned posture, and with active internal muscle chains is a dancer poised for the unexpected. Interestingly, individual precepts are indirectly benefical to establishing a dyadic resonance loop. Core-body tension and the right muscle chains pass on an

[2]Dynamic systems theorists might speak of *order parameters* that remain stable over time (as opposed to the more specific patterning of transient "movemes"). Order parameters designate collective variables that best represent the macroscopic patterns arising from non-linear interactions and synergies of lower-level components.

impulse received at the top to the feet without delay. When two people individually do this, a continuous inter-body chain connects A's legs via her torso to B's torso and down to his legs again. A precise axis also benefits the immediate registering of step onset for the partner, a sharp difference between action and inaction. A well-configured body thus furnishes information channels for the couple to use, without itself being a function of the dyad.

As chief permanent individual parameter, dancers must stay *in axis*: They remain upright and well aligned at all times, thereby creating micro-muscular support in the torso, lengthening the body for easy weight transfer in steps, and collecting its mass around an idealized line. This allows effortless pivoting with almost zero resistance, because all weight collects above a point on the forefoot. Teachers use didactic metaphors such as "**imagine being suspended from a string**", "**move the head upwards while the 'tail' drops to the ground**", "**imagine a vertical rod through your body**", "**never break your body line**". There also is a series of auxiliary techniques: Core tension by "**moving like a heavy object**"; alignment of shoulders-hips-knees-balls of feet ("**pivot en bloc**") and a one sided muscle column ("**build a firm tower over the hip of the supporting leg**"). In terms of imagery, the axis is a idealized perpendicular line running from head to toes, which epitomizes the dancer's position as if compressed and coincides with the rotation's center, i.e. the line that remains perfectly still.

To biomechanically validate axis control we measured the forward and lateral tilt of the torso segment when rotating around the partner in a multi-step *giro*,[3] using the upper light blue marker triangle in Fig. 2a, while Fig. 2b exemplifies one dancer's tilt over time. Averaged over all dancers the forward tilt has a standard deviation or mere ±2.5° (mean/range: −142°/13.2°). Similarly, lateral torso tilt reached no more than ±2.6° (mean/range: 2.0°/13.4°). Hence, axis control remains within a 5° range.

Good dynamic form presupposes respecting the "body grammar" of the axis: E. g., in the circulating follower's pivot (between *giro* steps) a proper axis visibly creates rotational precision. The dancer is firmly on a foot which aligns with hip and shoulder in a "**column**". An aligned axis is sensorially confirmed by minimal floor friction and absence of centrifugal forces and wobbliness. The dance partner's axis status also can be ascertained by rotating her very slightly. Simple visual inspection of our data confirms non-wobbliness: The hip marker's (Fig. 3a) trajectory nearly fits into a perfect circle (Fig. 3b). Figure 3c superimposes subsequent pivots of one follower, bearing witness to these formal constraints. Moreover, the sharp bend from circular to linear motion indicates the precise separation of pivots and walking. One teacher dubbed tango a "**line-dot**" system. An important didactic idea is that, when dancers reach the new position with their torso, they must "**collect**" and consolidate themselves there (to be prepared for both steps and pivots).

[3]This is a figure in which the follower orbits around the leader, using side-, forward-, side- and backsteps in succession.

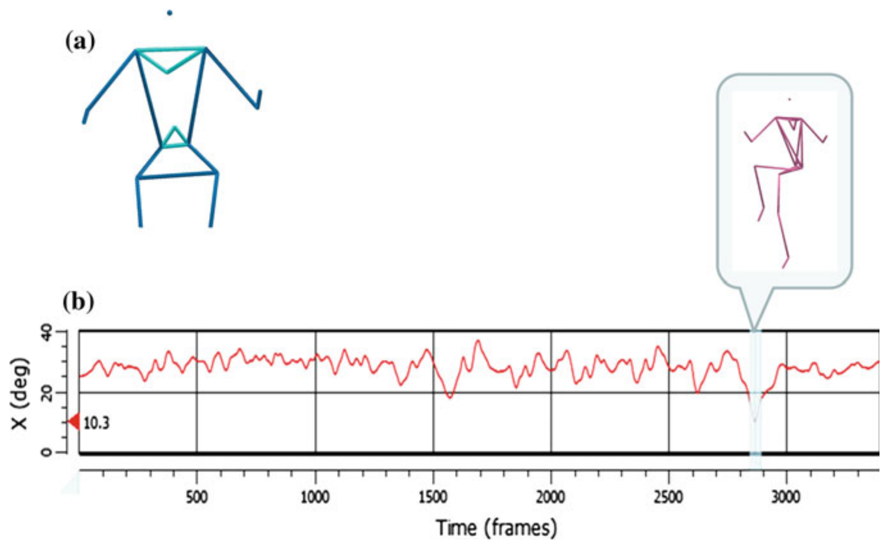


Fig. 2 Individual axis stability **a** Reference segments; **b** axis tilt over time; the blurb explains the outlier, a leg lifted for an ornamentation

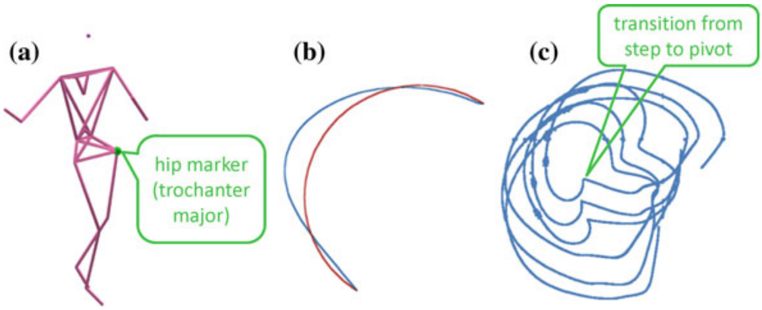


Fig. 3 **a** Marker placement, **b** trajectory (*blue*) inscribed into a perfect circle (*red*); **c** multiple successive pivots

2.2 Dyadic Habits

Complementarily, dyadic enabling properties are continuously activated. Dancers skillfully sense and attend to the partner, create spatial and tactile configurations for connectivity, configure the muscles to transmit signals, and establish joint anatomical structures. Dancers hereby aim at mutual interpenetration and an extension of the resonance loop into the partner's body. This allows them to dynamically incorporate his/her sensory feedback.

For connectivity and optimal force transmission both partners must notably point their breastbones—the "communication center"—towards each other, whatever the

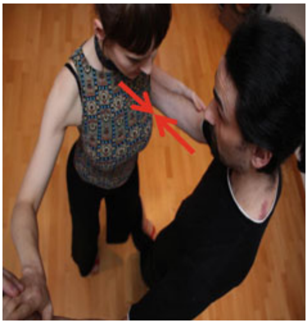


Fig. 4 Imagined breastbone vectors

hips and legs may be doing. Teachers often speak of imagining “**a magnet/rubber band or stick connecting the breastbones**”, “**shining a torch from one’s sternum at the partner’s sternum**”. Related didactic imagery calls for “**pointing the breastbones to an imaginary joint center hovering between the partners**” for positions of greater distension (Fig. 4).

A biomechanic comparison of shoulder girdles and hip segments validates this strong connection through the upper torso emphasized by teachers. We calculated orthogonal vectors emerging from both body fronts for purposes of illustration. When partners are fully opposite the upper and lower vectors naturally coincide, or almost (the couple in Fig. 5a has a standard embrace of 200°, not the perfect 180°).

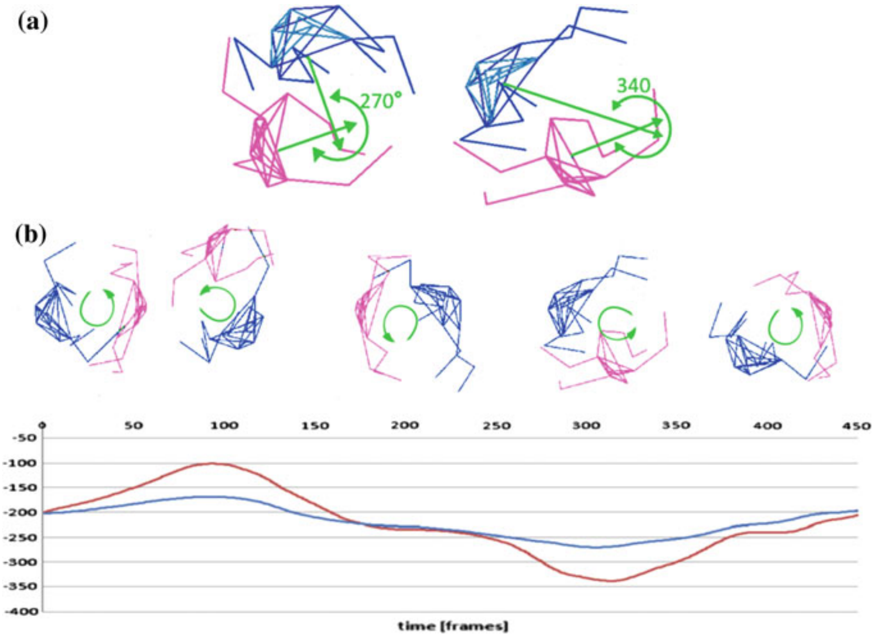


Fig. 5 **a** Angles between pelvis and sternum vectors in the couple in maximum twist (“dissociation”, viewed from *top*); **b** angles between both vectors over a step cycle

However, in some elements like in the *giro* or *ocho* the follower in orbit has to execute backsteps forcing her to twist her hips away from the partner. While this slightly affects the active communication center, the chest deviates far less than the hips when we compare the maxima of both angles (270° vs. 340°). To further clarify the point, Fig. 5b shows the differences of angles in a curve over a single rotation, with Δ-sternum appearing in blue and Δ-pelvis appearing in red. The pair configuration corresponding to each phase of the curve is shown at the top.

3 The Repertoire

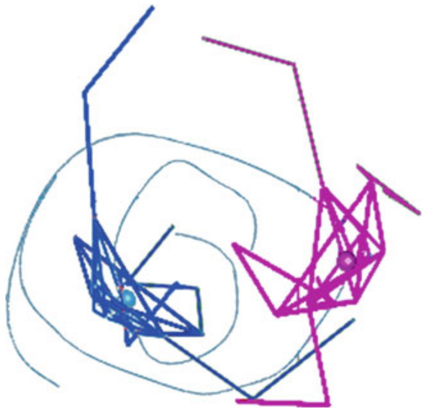
3.1 Geometries

An intuitive way of distinguishing typical multi-step tango patterns is to look at both dancers' trajectories based on their centers of gravity (COGs). Tango is rather geometric. Oblique steps are off-limits and cardinal directions "rule". Example, in the *giro* teachers often suggest imagery like "**create a square**" (or, depending on the teacher, "**a circle**"), "**pivot after each step**" and "**remain equidistant at all times**". Without being precise in this geometry the leader will be pulled off-axis while standing on one leg an executing this delicate maneuver. Kinematic data confirms the emergent dynamic geometry, in which the blue trajectories of both COGs approximate rounded squares nested within each other (Fig. 6).

Let us now compare different tango geometries. Figure 7 illustrates techniques performed by four couples:

- *Giro* (see above) with the leader in the rotational center and an orbiting follower.
- Partners together walk a rectangle in full opposite position, using forward-, side- and backsteps (*cuadrado*).
- The leader has the follower—whose front is turned about 90° away—do an "eight" by stepping back and forth on a line and pivoting at each end (*ochos*). (Note that dyad 3 compresses the zigzag movement onto a plane.)
- Both partners rotate around a common axis on a square trajectory with chests in opposition, while performing a succession of synchronized side-, back- and forward steps (4-point *giro*).
- The leader pivots the follower until her free leg extends and reverses this motion, creating a whip-like flying leg flourish (*boleo*).

Fig. 6 Floor geometry of a giro (*grey* trajectory and stick figures)



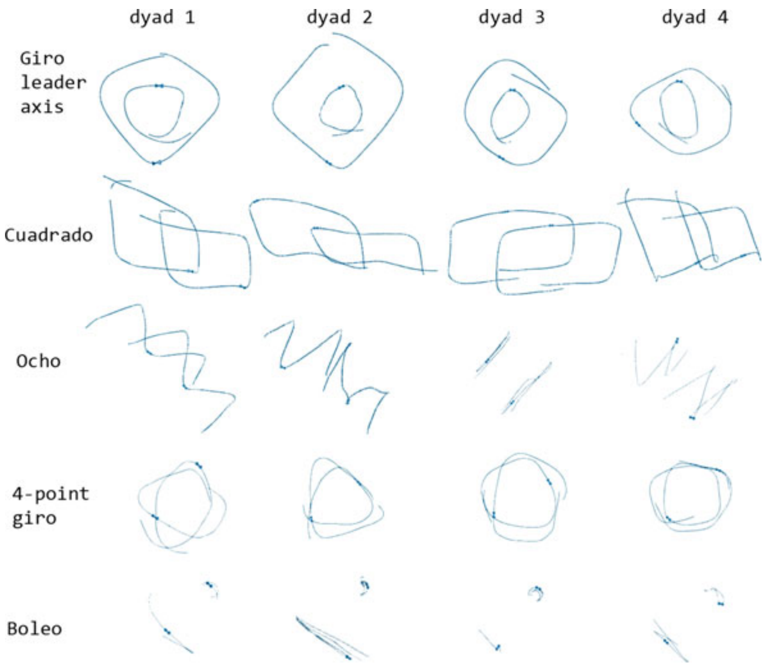


Fig. 7 Both dancers' floor geometry during different figures in four couples (COG-trajectories without stick figures, from *above*)

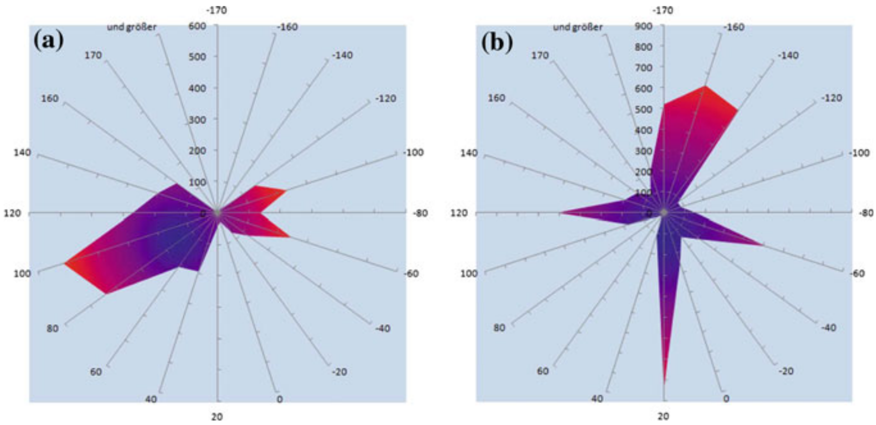


Fig. 8 Cumulative number of times a particular angle between walking direction and hip orientation occurs during a *giro* (**a**) and *cuadrado* (**b**) sequence, respectively. Spike length indicates the number of times a particular 10° sector was occupied over a multi-second sequence

3.2 Attractor Regimes

Visualization of tango elements can also be done through *state space* graphs that depict its "attractor regime", as termed by dynamic systems theory [18]. We chart what states occur how often, as differentiated from states remaining absent from the technique (or not occurring in tango at all). Figure 8 exemplifies this with one possible metric of many: the follower's twisting angle of her torso away from her walking direction. We compare a *giro* and *cuadrado* of one follower here.

4 Understanding Technique

Technical skills refer to how basic tango elements and at their transitions are efficiently executed (timescale 400–800 ms). This dynamic ordering of temporary tasks relates to what we called "vocabulary", i.e. the counterpart to permanent tango "grammar".

4.1 Individual Rotation

Let us inspect how dancers produce pivots, specifically, what the muscular "engine" for rotation is. Laypersons usually yank the body into a new orientation with active muscle power. Not so in tango where a release technique is used. The rotation is

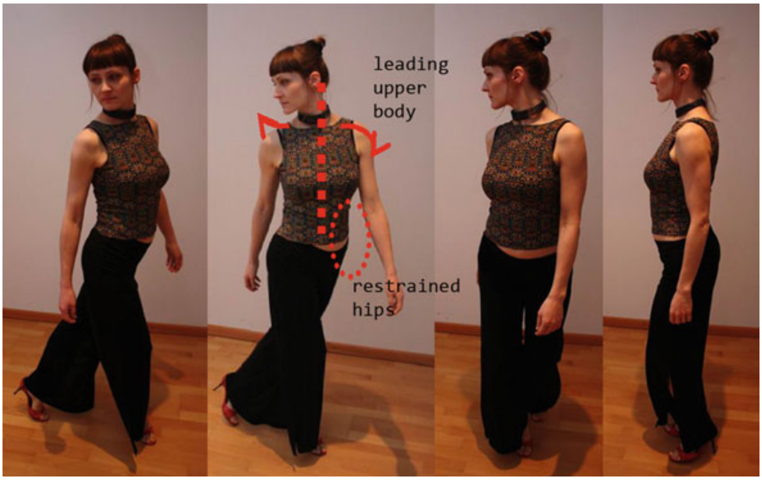


Fig. 9 “Dissociation” between upper and lower body to generate rotational energy

generated like a “**twisted wet towel unwinding**”. Initially, the twisting of chest to the desired new direction while the hips are actively restrained in place produces torsion. Provided that the abdominal and dorsal muscles connect the chest and the hips as part of proper general body organization, this “**loads up**” the body like a “**coil/spring**”. A brief moment of maximum “**dissociation**” occurs. The rotation then ensues by releasing tension. During recovery the hips catch up with the upper body’s new direction, as the dancer re-aligns (Fig. 9). All this produces a measured, but elastic and, if needed, powerful energy release for pivoting. A further “engine” comes from linear step energy that, upon re-arriving in axis and “**collecting**” the body, is converted into rotational energy.

To biomechanically validate this subjective kinetic model we calculated independent torso and pelvis vectors, as depicted in green in Fig. 10a. One vector emerges orthogonally at clavicle level and is defined against the plane of two shoulder plus one thoracic spine markers. The other vector emerges orthogonally from the plane spanning between both trochanters and the lower spinal marker.

We can now look at the vectors’ dynamic relation. A first measure done over several step-pivot cycles is the angular velocity, i.e. angle changes over time, of the follower’s upper torso (red) and pelvis (green) around the vertical axis (Fig. 10b). It details how, in the *giro*, the torso leads the movement and functions as a relatively stable *controlling segment*, whereas the pelvis qua *following/controlled segment* moves relative to it. Hence, the pelvis overshoots and catches up with the torso. Its relative oscillation around the more stable torso seems natural, as the hips need to be oriented in the exact stepping direction on an imaginary rectangle around the partner, while the torso stays oriented towards the center of the circle for communication with the leader.

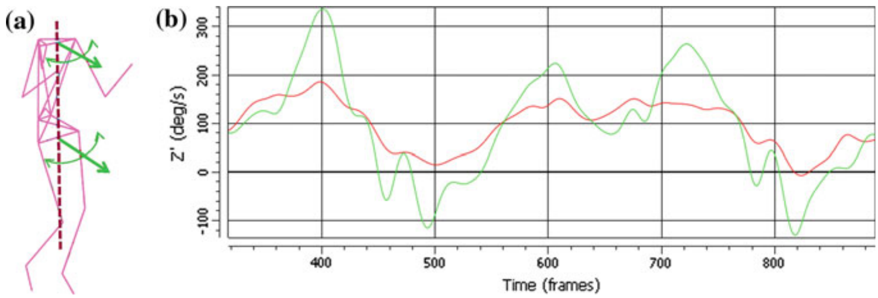


Fig. 10 **a** Marker placement and calculated vectors; **b** angular velocities of chest and hips over time

Next, we confirmed that individual pivots result from the above mentioned release technique by juxtaposing two graphs (Fig. 11): The moment the twisted upper body reaches maximum "**dissociation**" and begins to unwind back to its normal orientation precisely coincides with the line-to-curve transition of the COG on its spatial trajectory. In the top graph, the moment the line trajectory changes into a rotation is marked with a blue dot and the corresponding moment in the bottom graph with a blue line, indicating that the actual pivot begins a fraction of a second after muscle release onset. (The minimal delay probably owes to the body's configuring the muscles for transmitting the impulse downward.)

In sum, kinematic data frequently cannot access directly what the muscles do, etc. Clever ways of contrasting different data sets help make the appropriate inferences.

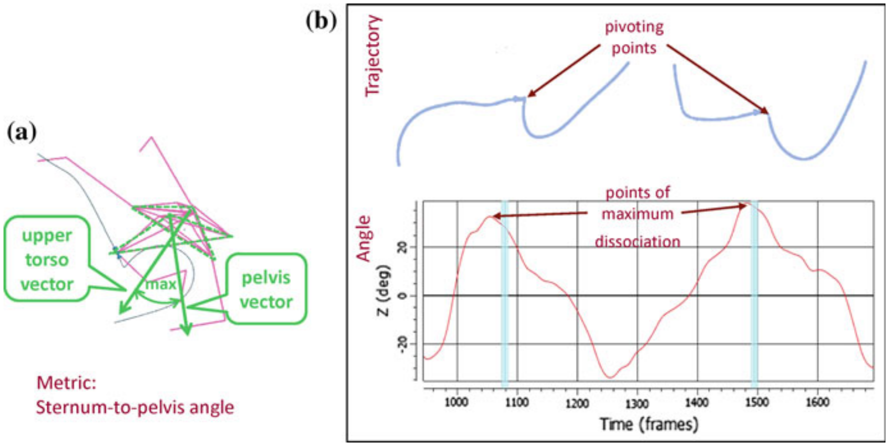


Fig. 11 **a** Marker placement and calculated vector; **b** Trajectory [*on top*] mapped on torso dissociation [*bottom*])

4.2 Dyadic Technique: Supported Rotation

Technique, of course, is frequently interpersonal. As a counterpart to the above kinetic model, let us elucidate the "engine" that is active when the follower is *passively* pivoted by the leader. The frequent scenario of a backward *ocho* (=eight) by the follower may serve as example. It begins with the leader putting the frontally opposed follower and himself on opposite legs and pivoting the follower by about 90 degrees. Then a joint lateral movement is executed, accomplished by the leader through a sidestep, whereas the follower does a backstep along a parallel trajectory. Once the follower collects her torso above her supporting leg again the leader gives her energy to pivot into the reverse direction. Now a step "backtracking" the trajectory can ensue, etc.

The subjective information flow at this juncture is this: The leader has occupied the extreme position on his trajectory before the follower gets there, senses the follower's footfall, weight on the leg and freeing of the other leg (a feel of "clicking into" a centered-weight position). Feeding forward energy into the follower earlier would irritate or topple her, as she cannot pivot before arriving in centered-weight position. The feedback about her degree of step completion signals to the leader the

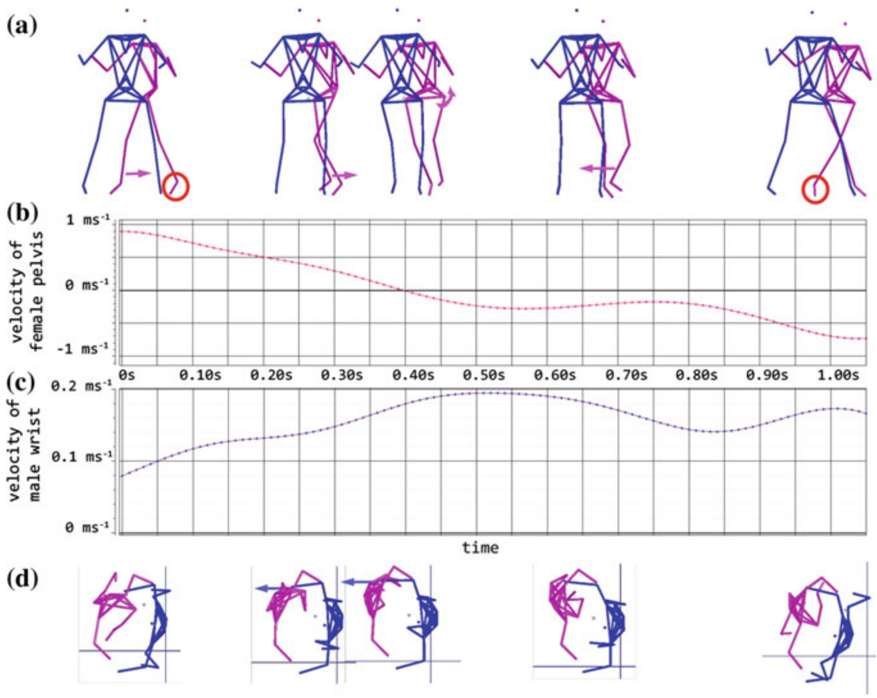


Fig. 12 **a** Stick figures of backward *ocho* with anchoring moments of follower's free leg *encircled*; **b** follower's pelvis velocity along the trajectory (note that the negative range of the graph depicts movement in opposite direction); **c** leader's wrist speed in his forward direction; **d** stick figures from above with wrist's energy direction depicted as *arrow*

onset-affordance for initiating the quick (guided) rotation. He now uses his torso's weight shift, arm action, and optionally the hip to exert peripheral force. Expressed biomechanically, the follower's foot can—at this moment—begin to act as fulcrum against which the leader can feed energy into the embrace via the arm-periphery. When this energy impacts the follower it makes her pivot. By swiveling away against the decentral force she can rapidly switch hip orientation by almost 180 degrees.

By measuring the leader's elbow-hand segment regarding its forward motion relative to the follower's translatory motion we were able to confirm this timing-dependent support biomechanically (Fig. 12): The leader's forward energy reaches its maximum the moment the follower's pelvis center comes to rest and her torso-hip "**dissociation**" neutralizes. Support for the rotation precisely coincides with the hips passing through a brief moment of opposition to the leader's body front just before rotating backwards into the next step again.

5 Understanding Interpersonal Coordination

Our multi-method approach can contribute to explaining how the supra-individual movement system is constituted with regard to micro-coordination. The question is how a dynamically stable joint entity emerges from two coordinated individual bodies, which in turn have their own substructure that needs to be coordinated. Since joint walking is frequently considered the supreme skill, our extended walking sub-study serves as an example.

5.1 Walking Together

As a first simplification we decided to summarize a single body's behavior regarding how the COG moves. This is justified to the extent that this nicely captures the the sum total of body weight in motion. Some dancers subjectively also think in such terms by imaginatively reducing their motion behavior to that of "**a ball situated behind their navel**" which rolls through space.

Utilizing the robust measure of COG velocity we investigated frontally opposed linear walking. The calculation was averaged over three steps from within each trial (steps 2–4, to eliminate acceleration/braking effects) and then averaged over six couples. For steps of all kinds we can straightforwardly distinguish the two subjectively most important step phases, mid-step and centered-weight when the torso passes over the supporting leg. Looking at leaders and followers individually (red and green lines, respectively), we see similar curves with small amplitude differences. This pattern is robust, with the surrounding lighter colored area representing the range of inter-step variation. Dancers *always* run through a sequence where the mid-step phase is fastest. This chimes with the idea that the centered-weight phase consolidates the dancer, even if one keeps "rolling on" (Fig. 13).

5.2 Tango Roles

Comparing the roles in the above figure, we discovered that, in 5 out of 6 couples, the follower's velocity curve slightly over-oscillated the leader's curve (The exception was a couple where this is not the case for a brief fraction of a second). This pattern, which may be general to asymmetric role dynamics [27], underwrites the leader's function as a dynamic anchor for the follower—after all, leaders minimally project ahead motor plans whereas followers respond in real-time. The pattern is one where followers necessarily reduce momentum more in the centered-weight consolidation phase (since they cannot know if the leader will continue with another step and since immobility is always preferable when in doubt), while leaders may "roll on", forcing followers to catch up by accelerating more. The subtleties of COG coordination thus point to the causality of a follower adaptively moving *relative to* the lead.

A comparable pattern is found in the *giro*. Here we compared both dancers' breastbones ("communication center") over time, which reflect the same pattern of the follower oscillating around the leader's velocity curve (Fig. 14). As in linear walking the follower cannot anticipate. Here, however, the challenge is exacerbated by moving on a longer peripheral trajectory and having to cover more space relative to the leader's shorter trajectory. For rotation scenarios like this, teachers emphasize that followers should "**follow the vector emerging from the leader's breastbone**" and "**occupy the emptied space**". Thus, the leader's chest rotating away from the opposed configuration signals when to initiate the circling and how long to continue it.

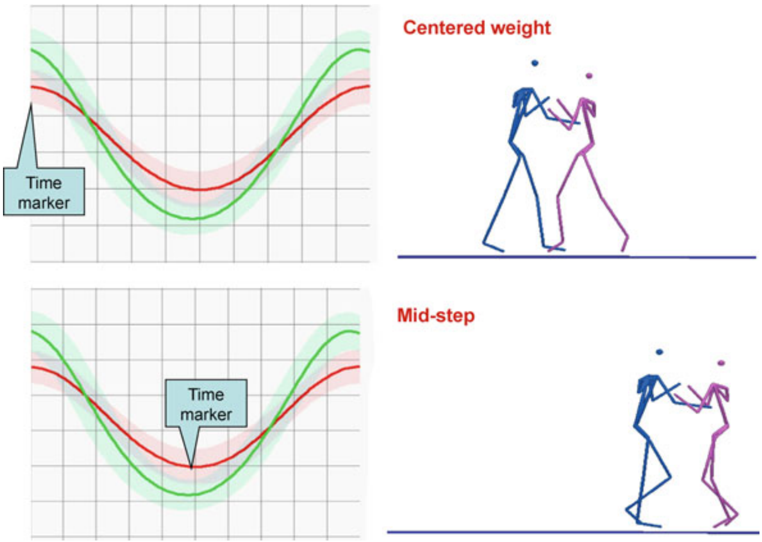


Fig. 13 Role-specific COG coordination dynamic of forward step cycle with two keyframes (metric: velocity)

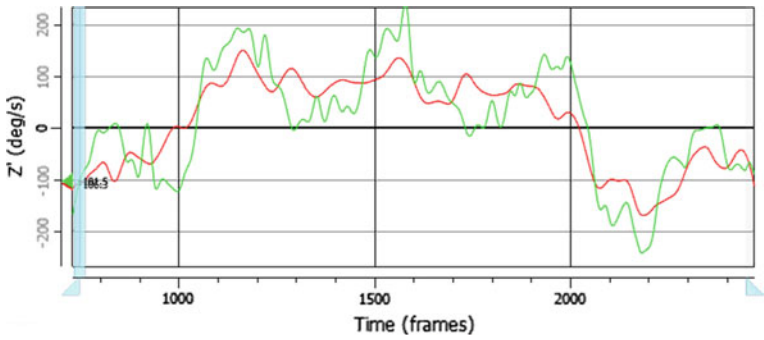


Fig. 14 Role asymmetry evidenced by chest velocities of leader (*red*) and follower (*green*) in the *giro*

The opening gap pulls the follower on the trajectory. This principle dynamically extends the breastbone opposition discussed earlier: When the leader leaves the opposed configuration the follower will immediately close the gap.

5.3 Coordination Quality

Dancers describe good coupling between partners as followers reacting to the lead immediately and with the just right increment of locomotion. As sensory correlates, dancers mention the feel of precise onset of the follower's self-motion in a step cycle and the joint weight shift of the body centers, as indicated by minimal pushing or pulling forces in the embrace. Nor should the follower overshoot at any time, which would be felt as putting pressure on the leader's embracing arm.

Biomechanically, good coupling should be reflected in a follower's minimal response delay, e.g. in straight walking. To measure how long after the leader the follower reaches the same keyframes of a step-cycle, as indicated by speed maxima and minima, we compared the pelvis velocities of leader and follower (Fig. 15). In effect, the time-difference of the peaks in each step-cycle of about 800 ms remain minimal: approximately 30 ms at centered-weight and 20 ms at mid-step position (standard deviation across step-cycle: 50 ms).

This delay is uncannily short. In fact, one might question if yet shorter would necessarily be better. According to what some teachers say minimal step-phasing delay may be desirable to keep the energy of the couple "rolling" and avoid the impression of the couple jointly "leaping" from one centered-weight position to the next.

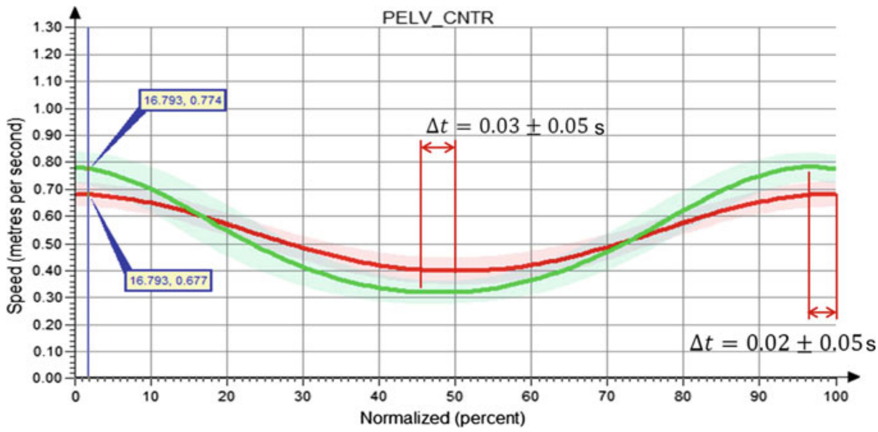


Fig. 15 Follower's delay in step-cycle at velocity maximum and minimum

5.4 *Coordination and Style*

We found tentative evidence which connects specific gait modulation patterns to different dance styles. For illustration, compare two couples: One couple teaches a style in which the follower's torso comes even closer to the leader ("**bread and jam**") before the actual weight shifts ensues, so a slightly fluid embrace with a "**controlled falling forward**" and relatively free muscle flow are used (Fig. 16a). The other couple emphasizes a style with a more rigid interpersonal connection (Fig. 16b). Greater muscular control through lower stance and more individual pushing of the quadriceps facilitates this coordinative style. The second couple's connective rigidity is reflected by the smaller inter-role amplitude difference, whereas in the first couple the follower needs to catch up slightly more mid-step and correspondingly slows down slightly more in the centered-weight position

5.5 *Complex Micro-coordination: Multi-body Part Contingencies*

As specified earlier, we intend to explain the collective dynamic by specifying what merges both individual dynamics into a well-structured emergent pattern. How do individual contributions give rise to phase-by-phase action interdependencies, from which coordination arises? To do so we must closely inspect how each phase of a technique is brought about at the timescale of a second or less. We propose to combine a subjective informational model (of active sensing strategies for perceptual triggers of micro-actions) with a behavioral model of observable time-locked patterns.

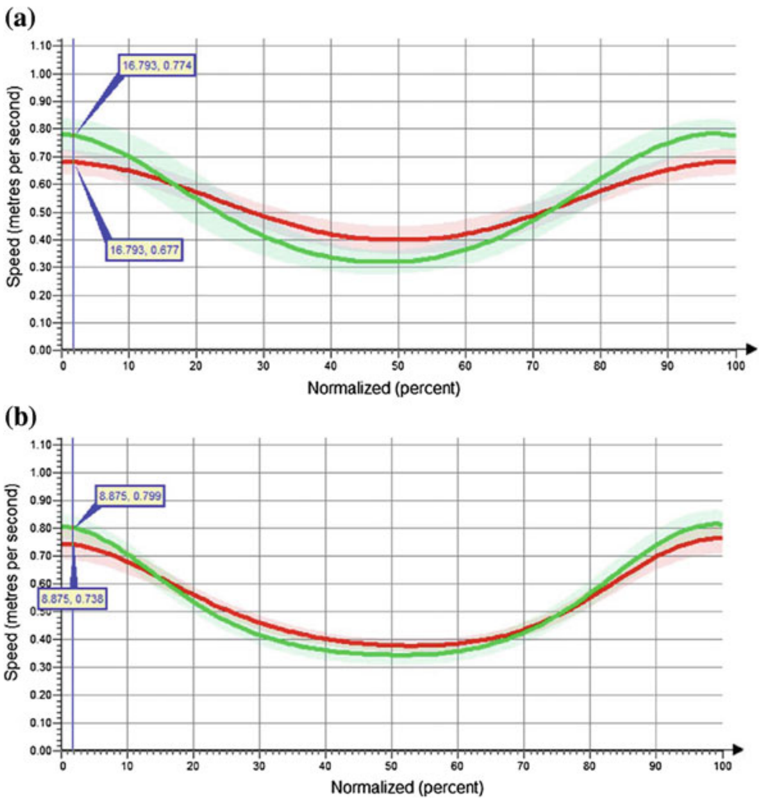


Fig. 16 Coordination dynamic of COGs in two couples: **a** variable connection, **b** more rigid connection

A single step's internal micro-coordination is structurally complex and accordingly challenging to model. While the COG accelerations of both dancers are clearly phase-locked, the legs are not, as shown in Fig. 17, nor are movement patterns across the two roles identical in any other way. The relatively straightforward in-phase relation of body centers is required for swift communication, being together, and (phasic) kinetic transmission from leader to follower, while the footwork contributes subservient synergies that are considerably more syncopated and inverted relative to where the body is facing – thus also more difficult to interpret.

When we model multiple body regions across the couple relative to each other this escapes any simple analysis. To make sense of such complexities, we need an informational model (cf. [24]) that explains how dancers coordinate their

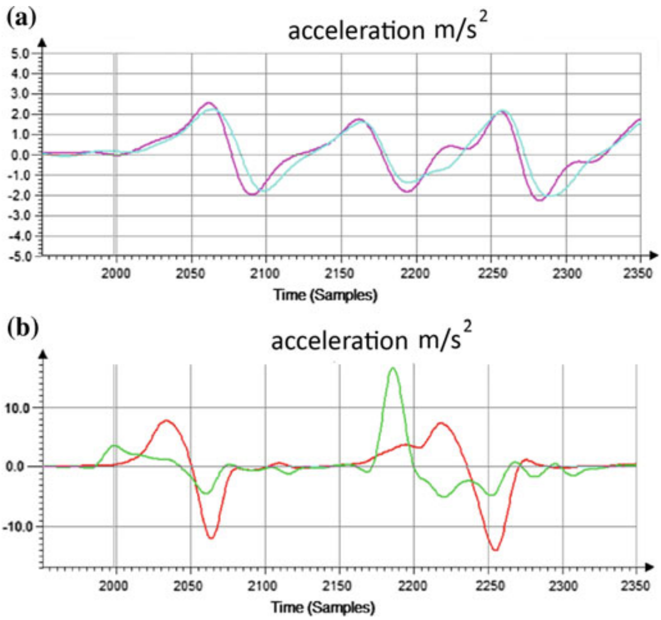


Fig. 17 Interpersonal coordination patterns of different body part accelerations: **a** body centers (COGs) in phase; **b** peripheries (here: ankles) out of phase

micro-actions with their partner. We reconstructed this from how some teachers actually train their pupils by systematically charting micro-percepts and micro-actions in both roles on a timeline. For illustration, let us stick with forward walking (as backsteps follow a somewhat different coordinative logic with different relative action onsets). The requirements of a joint weight transfer and the torso's staying together already have been discussed. Moreover, leg action needs to be coordinated to avoid collisions. The follower's free leg must have moved backwards before the leader sets his foot down in the same track. Also the step must end by collecting one's energy and without overshooting. Both partners briefly balance into a dynamic equilibrium, yet dynamically enough to keep the steps "rolling" if desired. The most powerful functional understanding of walking comes from the informational model we reconstructed based on ethnographic data from M. Kimmel's tango experience (Fig. 18):

(0) {Leader senses that both partners are in start position}
(1) {Forward weight projection of leader, a controlled lability, informs the follower of incipient intention}
(2) {follower "loads up" torso, but stays poised, while channeling the information to the free leg, which begins to extend backwards}
(3) {sensing the follower's increasing torso-leg stretch, the leader releases his weight fully while the follower activates supporting leg to generate thrust}.

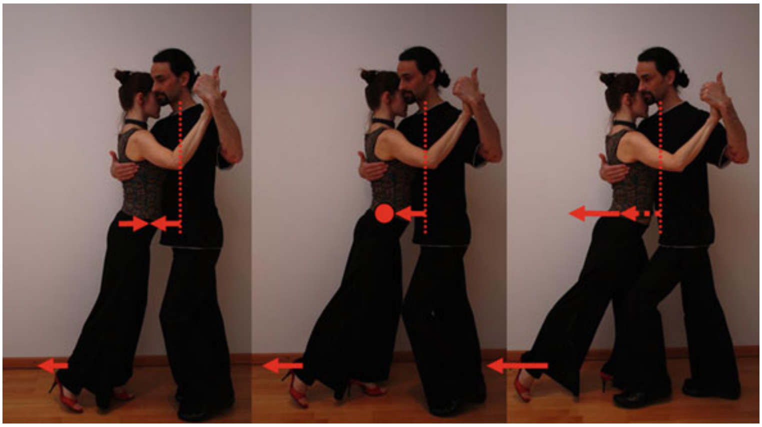


Fig. 18 Micro-coordination in a tango forward step expressed as subjective functional patterns (*arrows*, *axes*)

Before starting, the leader perceives action-readiness when both bodies are in torso-above-leg-support position using proprioceptive, balance and pressure signals. He begins to "**project**" the torso forward by a few centimeters into a position of controlled lability. The follower recognizes this incipient weight shift as slight incoming pressure and via diagonal forces in the embrace. She now responds in the free leg by beginning to extend it. This, in turn, can be felt by the leader through a slight pull at her shoulder-blade indicating a diagonally active muscle chain. When the leg has moved far enough, the leader can release more weight forward and again the follower responds by extending the step further. After her foot "**casts anchor**" she releases the other hip and leg, with the extended leg becoming the sole fulcrum. Now she may begin to actively push herself off while receiving the leader's body weight until both dancers pass through the start position again. The leader hereupon decides either to preserve the slight forward lability of his torso relative to his legs to "roll" over the relatively stable start position and into a new step or to make an extremely slight backward balancing maneuver to cancel the lability and jointly come to rest.

This micro-genetic reconstruction connects perceptual signatures dancer's use for orientation with the micro-actions these trigger. The resonance loops allows dancers to take guidance from active perceptual exploration of the partner (cf. [44]) and the dyad's configuration to mutually negotiate micro-actions phase-by-phase. (Such minute control is realistic, since at least in slow dancing, an initiated step can be paused or reversed.) The coordinated unfolding of a well-connected step emerges from the sum total of role-specific contributions. Importantly, it emerges from a set of causal interdependencies/contingencies stretching across multiple locales in both bodies: The follower's body part X reacts with action x_1 to action y_2 of body part Y of the leader, and vice versa, in a loop. This analysis summarily captures the complex causality of in-phase body centers and the more intricately

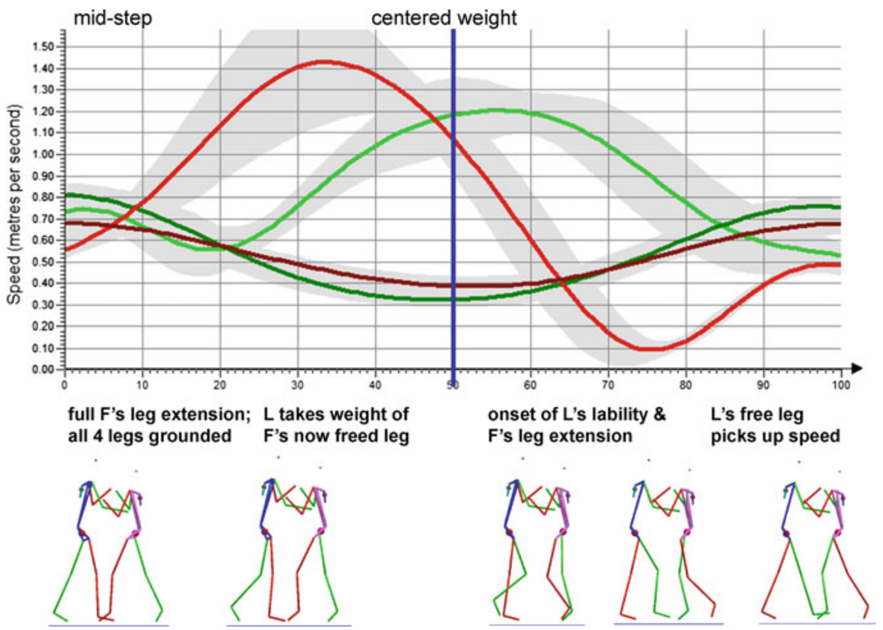


Fig. 19 Micro-coordination in a forward step at "cruising speed" expressed as velocities (*grey shading* standard deviation across six couples). *F* follower; *L* leader

coordinated leg substrates. Clearly, interaction control must be thought of as highly phase-specific in coordinatively complex joint locomotion.

The coordinative pattern can be further illuminated by biomechanics. Figure 19 depicts the velocity curves of both partner's COGs and one knee over a midstep-to-midstep cycle while "cruising" (averages exclude first steps). Right after the middle of the curve—the new step onset—the follower immediately accelerates her free leg whereas the leader waits before even minimally accelerating his own until a third of the step is over. Then he picks up speed while she remains moving. In terms of causality, a relative small incremental change of the leader's COG seems to bring about a relatively large change in the follower's leg extension as the step begins. Of course, to adduce this causality we must already understand the subjective functional logic of the lead.

In sum, our understanding of micro-coordinative causality of tango techniques—how the individual contributions produce the overall pattern—benefits from a subjectively derived model capturing the phasic cue structure, with micro-actions based on micro-triggers that are necessarily multisensory and thus unlikely to be easily captured kinematically. This informational model crucially supports the interpretation of biomechanic data. Future studies using multiple sensors could translate the informational analysis into role- and task-specific control laws used by dancers [50], i.e. algorithmic rules defining the required action increment in response to a perceptual increment like increased lability or a diagonal pulling forces.

6 Conclusion

Measurable dynamic order and the subjectively perceived informational coupling of dancers that it emerges from comprise flipsides of a coin. In analyzing complex interpersonal skills, combining measurements with subjective data therefore has much to commend it. Our present research "frontloaded" subjective data into a motion capture design, eventually engendering a genuine dialog. Both methods have specific strengths: Subjective data is crucial for directing the researcher's attention to functional principles, especially since the salience of movements (based on visible size, etc.,) or other things non-dancers might find striking, frequently misrepresent what really happens. What multi-sensory information has signal value for dancers is best explored subjectively. Conversely, kinematic measures best reveal subtle details of timing like relative movement onsets and micro-variations unlikely to be noticed such as the minimal velocity amplitude offsets of leaders and followers. Additionally, trajectories and related aspects can be visualized with intuitive conviction.

Furthermore, the necessity of drawing counterpart connections between two data pools proved highly instructive. It brought together the best of both worlds, forging a rich model of collaborative dance. Our method mix is especially apt for describing interpersonal coordination within tango elements, while also explaining how it emerges from both participants' micro-coordinated contributions. The causality of interaction can be thus revealed. Our methods contribute to a wider theory of participatory sense-making that illuminates how role-specific adaptive behaviors are dynamically linked—important steps to inspire control law based simulations, automatic parsing of biomechanic data, or even interactive robots. We look forward to this happening, as well as to mixed methods being applied to further collaborative dances.

Acknowledgments This research was supported by FWF grants P-19436 and P-23067. We wish to thank Mario Heller and Arnold Baca for assistance with organizing and setting up the motion capture study. We also thank the six teacher couples who participated in the motion capture study and our numerous interviewees. Finally, we express our gratitude towards Germano Milite and Betka Fislova for the photographic stills and for their generous support throughout the study.

References

1. D.K. Arvind, A. Valtazanos, Speckled tango dancers: real-time motion capture of two-body interactions using on-body wireless sensor networks. IEEE Comput. Soc., pp. 312–317 (2009)
2. A.L. Berkowitz, *The Improvising Mind. Cognition and Creativity in the Musical Moment* (Oxford UP, Oxford, 2010)
3. N. Bernstein, *The coordination and regulation of movements* (Pergamon, London, 1967)
4. J. Bourbousson, C. Sève, T. McGarry, Space–time coordination dynamics in basketball: Part 2. The interaction between the two teams. J. Sports Sci. **28**, 349–358 (2010)
5. M. Castro, *Tango Awareness* (Buenos Aires, self-published, 2004)

6. A. Clark, *Supersizing the Mind. Embodiment, Action, and Cognitive Extension* (Oxford UP, Oxford, 2008)
7. R. Dale, R. Fusaroli, N.D. Duran, D.C. Richardson, The self-organization of human interaction, in *The Psychology of Learning and Motivation*, ed. by B.H. Ross (Elsevier Inc., Academic Press, 2014), pp. 43–96
8. V. Fantasia, H. De Jaegher, A. Fasulo, We can work it out: an enactive look at cooperation. Front. Psychol. **5**(874), 1–11 (2014)
9. A. Fogel, *Developing Through Relationships. Origins of Communication, Self, and Culture* (The University of Chicago Press, Chicago, 1993)
10. A. Fogel, A. Garvey, H.-C. Hsu, D. West-Stroming, *Change Processes in Relationships: A Relational-Historical Research Approach* (Cambridge UP, Cambridge, 2006)
11. T. Froese, T. Fuchs, The extended body: a case study in the neurophenomenology of social interaction. Phenomenol. Cognit. Sci. **11**, 205–235 (2012)
12. T. Froese, C. Gould, A. Barrett, Re-viewing from within. A commentary on first- and second-person methods in the science of consciousness. Constr. Found. **6**, 254–269 (2011)
13. T. Fuchs, H. DeJaegher, Enactive intersubjectivity: participatory sense-making and mutual incorporation. Phenomenol. Cognit. Sci. **8**, 465–486 (2009)
14. S. Gallagher, Phenomenology and experimental design. J. Conscious. Stud. **10**, 85–99 (2003)
15. J.J. Gibson, *The Ecological Approach to Visual Perception* (Houghton Mifflin, Boston, 1979)
16. H. Haken, G. Schiepek, *Synergetik in der Psychologie: Selbstorganisation verstehen und gestalten* (Hofgrefe, Göttingen, 2006)
17. J. Ibanez-Gijon, A. Diaz, L.M.D. Lobo, On the ecological approach to information and control for roboticists. Int. J. Adv. Rob. Syst. **10**(265), 1–11 (2013)
18. S. Kelso, *Dynamic patterns* (MIT Press, Cambridge, MA, 1995)
19. M. Kimmel, Intersubjectivity at close quarters: shared, distributed, and superindividual imagery in tango argentino. Cognit. Semiot. **4**, 76–124 (2012)
20. M. Kimmel, A cognitive theory of joint improvisation: the case of tango argentino, in *Handbook of Dance Improvisation*, ed. by V. Midgelow (Oxford UP, Oxford, in press)
21. M. Kimmel, Sensorimotor and interaction management skills in tango argentino, in *Out for a Walk*, ed. by F. Engel, S. Marienberg, P. Schneider. Das Entgegenkommende Denken (Berlin, in press)
22. M. Kimmel, C. Irran, M. Luger, Bodywork as systemic and inter-enactive competence: Participatory process management in Feldenkrais Method® & Zen Shiatsu. Front. Psychol. **5** (1424), 1–23 (2015)
23. J. Lagarde, C. Peham, T. Licka, S. Kelso, Coordination dynamics of the horse-rider system. J. Mot. Behav. **37**, 418–424 (2005)
24. M. Lavelli, A.P. Pantoja, H. Hsu, D. Messinger, A. Fogel, Using microgenetic designs to study change processes, in *Handbook of Research Methods in Developmental Science*, ed. by D. Teti (Wiley-Blackwell, Malden, MA, 2006) pp. 40–65
25. M. Maduell, A. Wing, The dynamics of ensemble: the case for flamenco. Psychol. Music **35**, 591–627 (2007)
26. B. Magerko, W. Manzoul, M. Riedle, A. Baumer, D. Fuller, K. Luther, C. Pearce, An empirical study of cognition and theatrical improvisation, in *Proceedings of the Seventh ACM Conference on Creativity and Cognition*, pp. 117–126 (ACM, New York, 2009)
27. L. Noy, E. Dekel, U. Alon, The mirror game as a paradigm for studying the dynamics of two people improvising motion together. Proc. Natl. Acad. Sci. **108**, 20947–20952 (2011)
28. B. Olszewski, El Cuerpo del Baile: the kinetic and social fundaments of tango. Body Soc. **14** (2), 62–81 (2008)
29. O. Oullier, S. Kelso, Social coordination from the perspective of coordination dynamics, in *Encyclopedia of Complexity and Systems Sciences*, ed. by R.A. Meyers (Springer, Berlin, 2009), pp. 8198–8212
30. P. Passos, D. Araújo, K. Davids, L. Gouveia, S. Serpa, Interpersonal dynamics in sport: The role of artificial neural networks and 3-D analysis. Behav. Res. Methods **38**, 683–691 (2006)

31. C. Peham, T. Licka, H. Schobesberger, E. Meschan, Influence of the rider on the variability of the equine gait. Hum. Mov. Sci. **23**, 663–671 (2004)
32. C. Petitmengin, Describing one's subjective experience in the second person: an interview method for the science of conciousness. Phenomenol. Cognit. Sci. **5**(3–4), 229–269 (2006)
33. J. Rączaszek-Leonardi, I. Nomikou, K.J. Rohlfing, Young children's dialogical actions: The beginnings of purposeful intersubjectivity. IEEE Trans. Autonomous Mental Dev. **5**, 210–221 (2013)
34. M.A. Riley, M.J. Richardson, K. Shockley, V.C. Ramenzoni, Interpersonal synergies. Front. Psychol. **2**(38), 1–7 (2011)
35. K. Sawyer, *Group Creativity. Music, Theater, Collaboration*. Erlbaum, Mahwah, NJ (2003)
36. K. Sawyer, Individual and group creativity, in *The Cambridge Handbook of Creativity*, ed. by J.C. Kaufman, R.J. Sternberg (Cambridge UP, Cambridge, 2012), pp. 366–394
37. T. Schack. Building blocks and architecture of dance, in *The Neurocognition of Dance. Mind, Movement and Motor Skills*, ed. by B. Bläsing, M. Puttke, T. Schack (Psychology Press, Hove & New York, 2010) pp. 11–40
38. B. Schögler, Studying temporal co-ordination in jazz duets. Musicae scientiae **3**(1), suppl 75–91 (1999)
39. F. Seddon, M. Biasutti, A comparison of modes of communication between members of a string quartet and a jazz sextet. Psychol. Music **37**, 395–415 (2009)
40. D.N. Stern, *The interpersonal world of the infant: a view from psychoanalysis and developmental psychology* (Basic Books, New York, 1985)
41. R.H. Stevens, J.C. Gorman, Mapping cognitive attractors onto the dynamic landscapes of teamwork. Lect. Notes Comput. Sci. **678**, 366–375 (2011)
42. L. Tateo, The dialogical dance: self, identity construction, positioning and embodiment in tango dancers. Integr. Psychol. Behav. Sci. **48**, 299–321 (2014)
43. S. Torrance, T. Froese, An inter-enactive approach to agency: participatory sense-making, dynamics, and sociality. Humana Mente **15**, 21–53 (2011)
44. M.T. Turvey, C. Carello, Obtaining information by dynamic (effortful) touching. Philos. Trans. Royal Soc. B: Biol. Sci. **366**, 3123–3132 (2011)
45. F. van Alphen, Tango and enactivism: first steps in exploring the dynamics and experience of interaction. Integr. Psychol. Behav. Sci. **48**, 322–331 (2014)
46. F.J. Varela, J. Shear, *The View from Within. First Person Approaches to the Study of Consciousness* (Imprint Academic, Bowling Green, 1999)
47. P. Vermersch, *L'Entretien d' explicitation en formation continue et initiale* (ESF, Paris, 1994)
48. L. Vilar, D. Araújo, K. Davids, V. Correia, P.T. Esteves, Spatial-temporal constraints on decision-making during shooting performance in the team sport of futsal. J. Sports Sci. **31**, 840–846 (2013)
49. A.E. Walton, M.J. Richardson, P. Langland-Hassan, A. Chemero, Improvisation and the self-organization of multiple musical bodies. Front. Psychol. **6**(313), 1–9 (2015)
50. W.H. Warren, The dynamics of perception and action. Psychol. Rev. **113**, 358–389 (2006)
51. D.M. Wolpert, K. Doya, M. Kawato, A unifying computational framework for motor control and social interaction. Philos. Trans. R. Soc. Lond. B Biol. Sci. **358**, 593–602 (2003)
52. C.M. Zeitz, Some concrete advantages of abstraction: how experts' representations facilitate reasoning, in *Expertise in Context: Human and Machine*, ed. by P. Feltovich, K.M. Ford, R. Hoffman (AAAI, MIT Press, Menlo Park, CA, 1997), pp. 43–65
53. P. Zukow-Goldring, Assisted imitation: affordances, effectivities, and the mirror system in early language development, in *Action to Language Via the Mirror Neuron System*, ed. by M. A. Arbib (Cambridge UP, Cambridge, MA, 2006), pp. 469–500

Abstractions for Design-by-Humans of Heterogeneous Behaviors

Amy LaViers, Lin Bai, Masoud Bashiri, Gerald Heddy and Yu Sheng

Abstract This paper will discuss the use of movement observation, taxonomy, and expert knowledge, as found in Laban/Bartenieff Movement Studies to facilitate the production of diverse robotic behaviors that are easy to construct for a wide range of people. A 'behavior' will be defined by a set of movement primitives that are scaled and sequenced differently for distinct behaviors. These methods will be applied to real robotic platforms, in the context of manufacturing, motivating the fundamental value of high-level abstractions that produce a wide array of behavior. In particular, the integration of methods for designing an alternate set of "knobs" for motion control, abstracting sequencing behaviors and interactions of teams, and controlling robots via a web-based platform will be outlined in a language accessible to an interdisciplinary audience.

1 Introduction

In order to fully capitalize on the body of knowledge that is human movement, technological tools must be developed to enable design-by-humans of robotic behaviors and movement programs. This development requires an in-road to how humans conceive of designing movement. It also requires a set of tools that can balance the flexibility required for the variety of tasks humans might conceive and the rigidity with which a robot program must specify desired behavior. This chapter will review, for an interdisciplinary audience, a set of embodied movement con-

A. LaViers (✉)
Mechanical Science and Engineering, University of Illinois at Urbana-Champaign, 1206 W. Green Street, Urbana, IL 61801, USA
e-mail: alaviers@illinois.edu

L. Bai
Electrical and Computer Engineering, University of Virginia, PO Box 400743, Charlottesville, VA 22904, USA

M. Bashiri · G. Heddy · Y. Sheng
Systems and Information Engineering, University of Virginia, PO Box 400747, Charlottesville, VA 22904, USA

J.-P. Laumond and N. Abe (eds.), *Dance Notations and Robot Motion*, Springer Tracts in Advanced Robotics 111,
DOI 10.1007/978-3-319-25739-6_11

cepts for describing diverse movement generation and previous technical work that aligns with these concepts. It will also present a new computational framework that leverages the prior, more fundamental technical work to realize the vision of a system which facilitates the creation of diverse, heterogeneous behaviors that do not require technical training to create.

Humans are constantly in motion. Every task in daily life requires movement, and many skilled jobs required highly specialized motion to accomplish. The line of research discussed in this chapter is meant to leverage the incomparable knowledge that humans—either individuals or in groups—contain about movement. Rather than modeling an entire robot behavior [1], our focus is on creating a tool such that people can craft movements on a sophisticated robot, without extensive technical training.

Robots move based on computational frameworks that are highly specialized. In scenarios where economies of scale allow for teams of engineers to implement and refine precise, repeatable movements, robots and automated systems have contributed a great deal to increasing production and easing the arduous physical tasks involved in high volume production lines. However, the process of reprogramming such behavior is time consuming and expensive, requiring teams of technically trained personnel. This barrier prevents line workers, small businesses, and lay individuals from instantiating such behaviors themselves.

Four key technical ideas will be leveraged in the framework—two from LBMS and two from robotics. On the part of sequencing, we will use *motif symbols* and *finite state machines*. Motif symbols (a derivative of Labanotation) will be used as flexible labelings of basic motion primitives. These motion primitives—and their compatibility at any given stage of a developed movement phrase—will be represented with a finite state machine. To allow for primitive modulation, or the idea that a single primitive may be executed in several different ways, we will use *Motion Factors* and *optimal control*. Motion Factors are enumerated in LBMS as a description of movement Effort; these factors will align with parameters in an optimal control problem such that the modulation of movement trajectories is informed by the experience of movement quality as given by Laban.

The system laid out by Rudolf Laban in the early 20th century codified terms and concepts and continues to influence how dancers describe their movement and how they train themselves to perform a greater range of movements. Today, this work is known as Laban/Bartenieff Movement Studies (LBMS) and remains an in-road to expert knowledge and embodied movement definition [2]. LBMS is an established method for movement analysis and observation with three institutes, the Laban/Bartenieff Institute for Movement Studies (LIMS) in New York City, Integrated Movement Studies (IMS) which operates out of the University of Utah, and Trinity Laban Conservatoire of Music and Dance in London, that offer certification programs in the work. The framework stems from an embodied movement perspective and is utilized in many professional contexts, such as therapy, consulting, and choreography. The corpus of the framework was largely established by Laban and his student Irmgard Bartenieff, as described in their work [2–7]. It is divided into four parts: Body, Space, Effort, and Shape; these include several overarching principles and dualities, size fundamental movements that can be used to retrain body patters, four parameters that govern movement quality, several key platonic solids that depict spatial movement options, a system, and shorthand, for movement notation, classifications of body shape, and modes of shape change. Our system will leverage elements in the Body and Effort categories.

Several previous technical publications [8–11] introduce a method for producing *stylistic* movement phrases in a principled way. These stylistic behaviors lend themselves to diverse *high-level* behavior production. We distinguish a high-level behavior as one that is characterized by the arrangement and execution of movements, rather than a single isolated action (which will be produced by a *low-level* controller). The method uses a notion of human movement, outlined in this section, that separates movement ordering from exact execution; this implicit segmentation of movement into distinct "moves" or *primary motions* is tailored for handling our notion of *style* in a quantitative way. That is, the order of moves and the way each move is executed defines a style of movement and, thus, these quantities should be allowed to vary independently. As such, the stylistic parameters fall into two categories: those which govern sequencing rules and those which specify each movement trajectory. A novel computational platform will unify several of these parameters at a high level for ease of use by an untrained operator.

The rest of this chapter is organized as follows: Sects. 2 and 3 contain key concepts from LBMS and robotics presented synchronously in order to draw important parallels that will be leveraged in the computational framework. Section 4 reviews prior, previously published work that will set the stage for the way of thinking in our novel system. Section 5.1 reviews existing computational frameworks that our system will leverage and replace with a system at a higher level of abstraction. Section 5.2 will describe our novel system. Section 6 will describe how the system can be leveraged in mechanizing movements found in industry and homes around the world. Finally, Sect. 7 summarizes the contribution of this chapter.

2 Motif Symbols and a Discrete Event Description of Movement

The first barrier to would-be movement designers encounter is that of establishing a movement program. Where people tend to describe actions using stereotyped vocabulary and embodied actions, i.e., verbs and demonstrations, robots require time-varying signals to each mechanism comprising their physical platform. Such signals are often controlled by digital logic design via small microprocessors or programmable logic chips (PLCS); both of these devices require a fully specified instruction set to execute. For example, if you ask a human to execute a series of three, possibly unrelated and unconnected, actions, the human mover will figure out a physically feasible action to connect the three movement. A machine cannot do this, and this section will lay out a description of the type of description it instead requires and outline a way to label these descriptions using motif symbols.

2.1 Finite State Machines

We will begin with how the behavior of a simple device, requiring movement from a human operator, is represented. This *discrete-event description* of behavior is how we will represent movement in our final system. Engineers find it useful to construct abstract machines that represent the discrete event behavior of the system called automata, or finite state machines, to better understand inputs, outputs, and the intermediate stages or states. Our example is a TV remote that allows a user to flip between three TV channels. This is the scenario painted in Fig. 1.

The user may press the channel '*up*' or '*down*' button on the remote as in Fig. 2. We call these options actions or *events*. This causes the TV to change its channel or *state*. "State" is a central concept in engineering—it allows for control of engineered systems by defining the fundamental essence of the system. In the case of automata, we have a *discrete state space*. For example, the state space of the system in Fig. 1 may be written as

$$X = \{01, 02, 03\}.$$

Similarly, the space of events is also discrete—as imaged in Fig. 2—and may be written as

$$E = \{up, down\}.$$

See Fig. 3 for a conceptual comparison of the concepts of discrete and continuous. Then, the structure in Fig. 1 specifies how the device moves between these states based on the actions of a user. This structure allows for engineers to translate this behavior right down to the circuits, the transistors, of the remote control (and TV).

Discrete event systems have three major components: a set of states (the circles in Fig. 1), a set of events, or transitions, (the arrows in Fig. 1), and an arrangement, given by a mapping function, of how these two connect (depicted in Fig. 1). The machines also typically name an initial state (indicated by the arrow terminating,

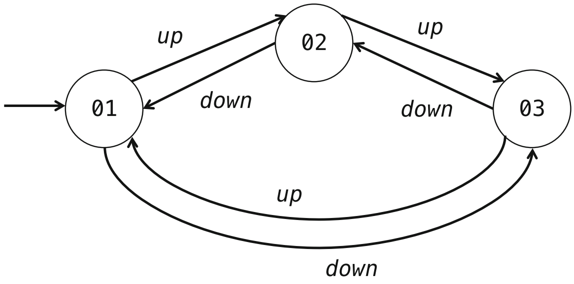


Fig. 1 The finite state machine governing the behavior of a basic remote control flipping between channels on a TV with three channels

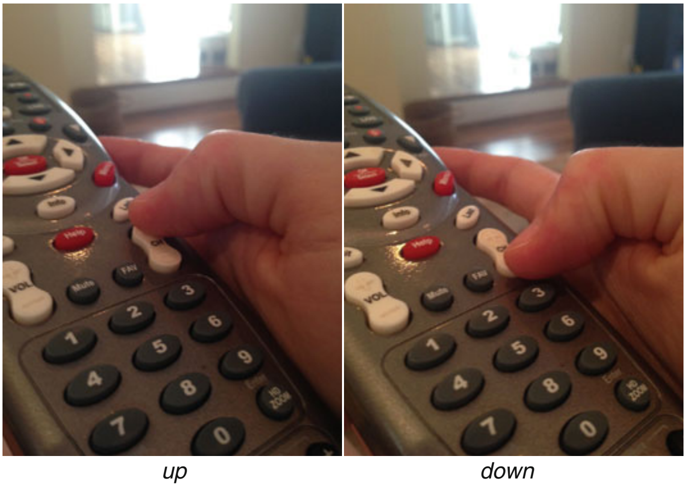


Fig. 2 The *movements* a user employs to use this device. These movements only occur in the specific order embedded underlying finite state machine

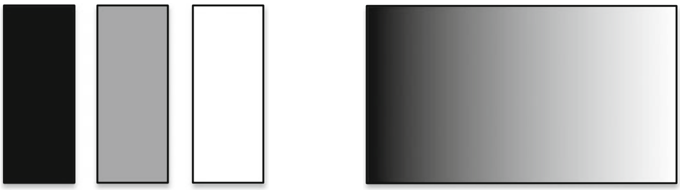


Fig. 3 A finite versus infinite representation of color gradation. In the image on the left, the fade from black to white is represented by three *discrete* objects. In the image on the right, this gradation is represented in a *continuous* way

but not originating, in a state in Fig. 1) and a "marked" state, which indicates when a sequence through the machine may terminate.

2.2 *Style-Based Movement Sequencing*

In a similar way, we will define the principal element of the movement model utilized in this system: a discrete event description of primary motions. These pre-defined motions are abstracted as events in a discrete event system—defined only by their starting and ending poses. Our system will deal with sequencing and modulating these motions once they are defined, but much work has gone into creating easy ways to define a single motion on a robot [12, 13]. Later, interpolations between these poses are used as reference signals for an optimal control

Table 1 A list of barre exercises [8] and their corresponding transition labels utilized in Fig. 4

Movement	Transition labels
plié	$plie_i$, $plie_o$
relevé	$rele_i$, $rele_o$
battement tendu	$tend_i$, $tend_o$
degajé	$dega_i$, $dega_o$
coupé	$coup_i$, $coup_o$
frappé	$frap_i$, $frap_o$
passé	$pass_i$, $pass_o$
développé	$deve_i$, $deve_o$
battement	$batt_i$, $batt_o$
grand battement	$gran_i$, $gran_o$

problem that can modulate exactly how each motion is executed. Thus, from the simple system first enumerated, structured combinations and modulations of these primary movements can be produced. The structure of this discrete event system will form the basis of the correctness of the robot program.

This stance is inspired by the training used in classical ballet; for example, the movements performed at the barre, classical ballet's warm-up exercise, make great candidate primary movements. These movements [14], listed in Table 1, connect ten basic body positions which form the states of our discrete event description of movement, as in Fig. 4 and prior work [8]. We distinguish two transitions for each movement listed in the table using a subscript to indicate an *in* and *out* variant. Thus, a movement is either going towards or away from each of the two states (body positions) it connects, and the next state depends on which of these is the case.

As in prior work [8] the arrangements of these movements maybe represented as a finite state machine, which will give the discrete ordering that is allowable of these basic, simple movements. This is essential for the system we present later, which needs a definition for allowable movement sequences to ensure that the operator gives feasible commands to the robot platform.

$$\mathscr{G} = (X, E, O, f, \Gamma, o, x_0, X_m, \varepsilon, \omega), \tag{1}$$

where X is the finite state space, E is the event set, O is an output set, $f\colon X \times E \to X$ is the state transition function, $\Gamma\colon X \to 2^E$ is the set of feasible events (at a given state), $o\colon X \times E \to O$ is the output map, $x_0 \in X$ is the initial condition, and $X_m \subseteq X$ is a set of marked states. In order to allow for both synchronous and asynchronous transitions in the Cartesian composition of two such systems, "empty" transitions, which are defined by the symbol ε, are used. The interpretation is that for the finite state machine, we insist on $\varepsilon \in E$, with the result that $\varepsilon \in \Gamma(x)$ as well as $f(x, \varepsilon) = x, \forall\, x \in X$. Moreover, $\omega \in O$ is associated with the outputs from "empty" events, i.e., $o(x, \varepsilon) = \omega, \forall\, x \in X$ and could be used in more complex systems for restriction of allowable sequences.

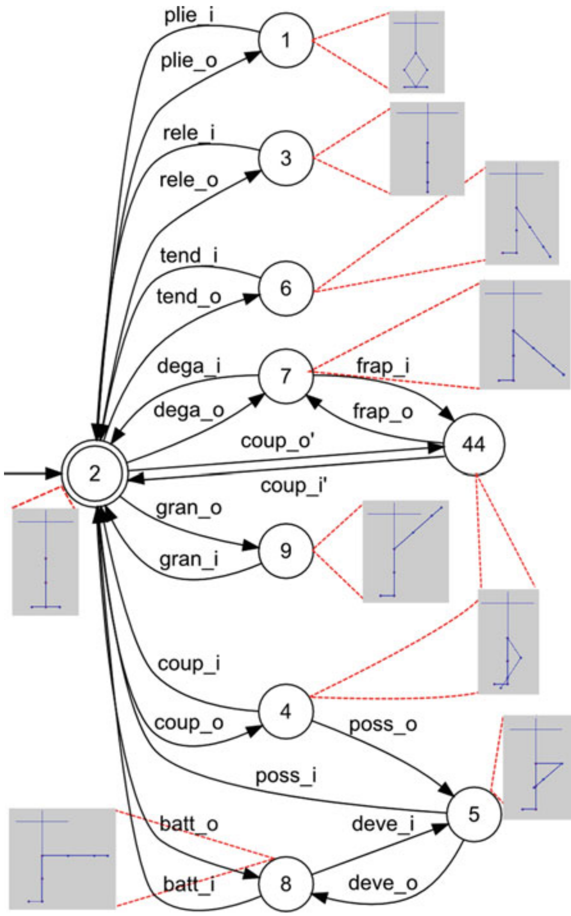


Fig. 4 A transition system which models the working leg (*right leg*) of a dancer during a ballet barre exercise. This figure illustrates the basic element of our movement model: a discrete event description of primary motions [8]

In order to demonstrate how a sample path through the system works, we give here the description given in our prior work [8]. Consider, for example, a développé; this movement is found both in barre exercises and more complex ballet movement phrases. A développé is the action when the working leg's foot is moved from the knee and then extends from the body. At the barre this motion is practiced by moving the working foot from standing to the ankle, then the knee, then extends from the body so that the leg is parallel to the floor. Lifting the foot to the ankle or knee (without, for example, any extension to follow) are allowable movements called coupé and possé, respectively. Thus, our uniquely defined transitions (and trajectories) are three separate events for the working leg: $coup_o$, $poss_o$, $deve_o$. Next, the dancer performs a closing movement where the foot remains extended from the body and the leg is lowered till the foot is returned to the starting stance. This is modeled as the event $batt_i$—the transition from pose 8 directly to pose 2 with a label that corresponds to the *in*-trajectory of a battement (a simpler movement that looks like a high straight-legged kick). The events $coup_i$, $poss_i$, and $deve_i$ are also defined, that is, the reverse pose transitions are allowed and used for more complex movements. And, a special event is enumerated to describe the inaction of the standing leg.

2.3 Motif

While ballet is highly formalized and contains stereotyped, named movements as above, the applications in which we may want to create a robot program may not be. Thus, we look to LBMS for a more flexible description of stereotyped movement. Motif is a method for capturing the essence of a movement sequence and is cataloged in work from the LBMS community [15]. This system allows people to figure out the main character of a motion rather than prescribing a precise, detailed pattern. Within the system are symbols which classify broad categories of actions. Six commonly used symbols that describe such *basic body actions* are given in Fig. 5.

The advantage of this notation system is that limb to limb or joint to joint motion recording is not necessary, only the essence of the motion is required. Thus, it could be a platform invariant method of notation, for example, a single arm robotic manipulator might have several joint angles and far different from the topology of human body. However, if the manipulator changes from 'C' shape to 'I' shape, it is extension—and the same classification can be given to a similar action on a human despite the fact that their arm produced a much more complex action. Our system will use these symbols to communicate to the user the nature of movements available for their use. It is a symbolic labeling that will easily communicate the nature of the discrete event system and action space of the robot platform (Fig. 6).

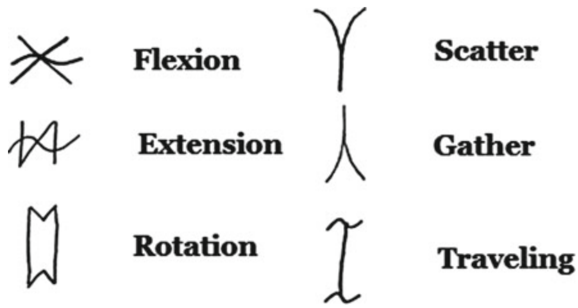


Fig. 5 Symbols from Motif description. Each *symbol* is aimed at getting at the essence of the movement. These are six commonly used symbols for *basic body actions*. This figure is recreated from work in the LBMS community [16]

Fig. 6 A motif of a movement phrase. The motif works up the page with time. *Symbols* may refer to the Shape, Space, or Effort as well as use the list of basic body actions in Fig. 2

3 Effort and Movement Modulation

The next layer of the framework allows for a given movement sequence to be executed in slightly different ways. This ability further increases the diversity and heterogeneity our system facilitates. This section reviews more of our prior work [11].

3.1 Effort and Our Mapping to a Cost Function

Laban names four categories of Effort or *Motion Factors*: *Space*, *Weight*, *Time*, and *Flow*. Space, Weight, and Time are interrelated via the structure in Fig. 7. In a given instance of a movement, each factor may take on one of two qualities. These qualities represent the extreme notions of each motion factor. These extremes of the first three Motion Factors combine pairwise to form the eight basic efforts: dabbing, gliding, floating, flicking, thrusting, pressing, wringing, and slashing. Each of these terms corresponds to a familiar pedestrian action, highlighting, even to a lay audience, the nature of the Dynamosphere arrangement: changing the quality of one motion factor moves around the cube to a different Basic Effort Action (BEA). These three motion factors, and the fourth factor, flow, which describes the quality of the connection between movements, are described—along with some intuition behind our mathematical interpretation.

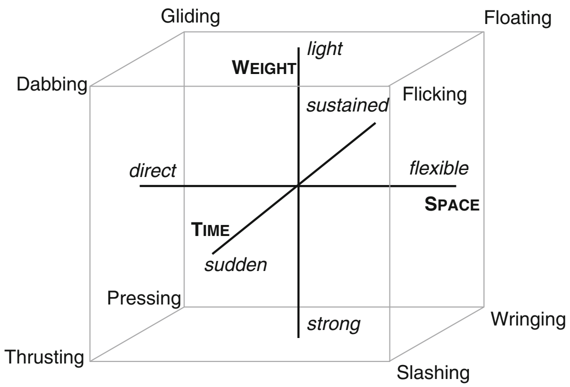


Fig. 7 The dynamosphere is Laban's arrangement of eight basic effort actions according to the axes of space, weight, and time. In **bold** are the three Laban *Motion Factors* which deal with single movements; in *italics* are the two qualities Laban associates with each factor; and in plain font are the eight basic efforts which result from the pairwise combination of each quality. Figure recreated from work in the LBMS community [7]

The Space axis describes how the dancer's attitude toward space is perceived. *Flexible* movements seem more carefree, meandering, and indirect; *direct* motions appear more matter of fact and judicious with their use of space, focused on arriving at a particular point. We pair this concept with a system's notion of reference tracking; direct movements track their path more aggressively than flexible ones. Thus, we will make use of nominal, linear trajectories away from which our solutions may deviate or adhere closely. The axis of Weight deals with the emanated sense of weight in the dancer's body during the movement. *Light* movements look as though they are less influenced by gravity—perhaps they are effortless—whereas *strong* movements are muscular and seem taxing on the body to perform. We interpret this as a specification for how much control effort (or input) is used to perform the movement. Thirdly, on the Time axis on a movement may either be *sudden* or *sustained*. This describes a quality which is more subtle than just the duration of the movement and a movement that lasts 5 s may be executed with a sudden or sustained quality. We interpret this as a metric over how much the state of the system is allowed to change along the trajectory of a movement: during a sustained movement it evolve more gradually while in a sudden movement it may deviate wildly in a single time step. Finally, movements' use of Flow may either be *free* or *bound*. In free flow a dancer seems to move through movements without the ability to stop their movement with any immediacy; while in bound flow, the dancer appears more careful to execute the succession of movements precisely. We interpret this from a systems perspective as a lesser (or greater) desire for the dancer to hit poses between movements exactly. In the sequencing framework we will employ, this translates to varying the weights on the desired terminal pose.

3.2 Style-Based Movement Modulation

In this section, we use a linear quadratic optimal control framework to find new time-varying trajectories between static poses through a mapping between Laban's effort system and the weights in our cost function that was first proposed in our prior work [8]. Our mapping may be considered an alternate mapping to a prior mapping [17] and later similar mappings [18, 19].

We first establish a linear system with an input $u \in \mathbb{R}^m$, a state $x \in \mathbb{R}^n$, and an output $y \in \mathbb{R}^l$ which tracks a reference signal $r \in \mathbb{R}^l$. This linear system will not approximate the dynamics of our robot platform; instead it operates at a higher level of abstraction where we are crafting a desired trajectory; low-level controllers will be presented later to execute this trajectory on our particular platform. We then establish a quadratic cost function

$$J = \frac{1}{2}\int_0^{T_f} \left[(y-r)^T Q(y-r) + u^T R u + \dot{x}^T P \dot{x}\right] \mathrm{d}t + \frac{1}{2}(y-r)^T S(y-r)\big|_{T_f} \quad (2)$$

in order to find an input u principled on the weight matrices $Q \in \mathbb{R}^{l\times l}$, $R \in \mathbb{R}^{m\times m}$, $P \in \mathbb{R}^{n\times n}$, and $S \in \mathbb{R}^{l\times l}$. The final time T_f represents the length of time each individual movement takes to execute. By construction, each of these matrices are positive definite and symmetric. Furthermore, their entries create a continuous, quantitative version of Laban's effort system and will determine which movement qualities are exhibited by the optimal trajectory, e.g., the trajectory may be bound, direct, sudden, and strong.

The relative weightings of the matrices will determine this, and weights in the middle may be seen as Motion Factors that are less present for this particular movement (as their weighting is more or less "neutral"—not expensive or cheap from the perspective of this cost function). Based on the discussion in the previous subsection, we associate the weight Q to the Laban's motion factor space, R with the factor weight, P with time, and S with flow. These weights correlate with the quality of each factor as follows:

$$Q \sim direct \quad (3)$$

$$R \sim light \quad (4)$$

$$P \sim sustained \quad (5)$$

$$S \sim bound \quad (6)$$

where the opposite of the qualities listed, flexible, strong, sudden, and free, are achieved when these weights are relatively small, respectively.

Using these weights as the style-based parameters for varying the resulting trajectory, we solve the optimal control problem

$$\begin{aligned} &\min_u J \\ &s.t. \\ &\dot{x} = Ax + Bu \\ &y = Cx \end{aligned} \tag{7}$$

where $A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times m}$, and $C \in \mathbb{R}^{l\times n}$. See our prior work [11] for a derivation of first-order necessary conditions for this dynamically constrained optimization. Importantly, we are able to develop algebraic conditions for the optimal state and costate which allow for movement classification via an inverse optimal control problem in later publications [20, 21].

4 An Illustrative Example for Robotics

Now we demonstrate the combined utility of this framework for specifying desired stylistic robotic behavior by constructing a disco dancing and a cheerleading behavior as presented in our previous work [11]. We produce behaviors that are perceived by viewers as being different, and thus, heterogeneous. For this section, we will use poses on the Aldebaran NAO robotic platform to demonstrate the basic idea behind the work. In the next section, we will develop a particular computational framework applied to Rethink Robotics' Baxter Research Robot.

First, we construct two instantiations of one-arm automata $\mathscr{G}^{disco}_{arm_{1,2}}$ and $\mathscr{G}^{cheer}_{arm_{1,2}}$. These objects assemble establishes allowable sequences of movements; in prior work [11] we show how more complex sequences can be enumerated with composition operators. Then, we enumerate two sets of weights $\{Q, R, P, S\}_{disco}$ and $\{Q, R, P, S\}_{cheer}$ which determine the effort quality of our trajectories between poses according to a cost function corresponding to the one in Eq. (2). Thus, we specify rules for motion sequencing and the dynamic timing each motion should exhibit.

Each state of the one-armed automata, which corresponds to a single arm pose, is constructed from a pair of joint angles, (θ_1, θ_2), as shown in Fig. 8. These poses were chosen such that the degrees of freedom of the arm (limited to the body's coronal plane) were discretized *by a would-be user* and are shown in Fig. 9. From the user's perspective, these shapes are deemed to be critical to the experience of behaviors such as cheerleading and disco-style dancing. From Fig. 8 and the general kinematic constraints of humanoid geometry, we see that the range of $\mathscr{X}$, for both behaviors, is limited, e.g.,

$$\mathscr{X} = \{(\theta_1, \theta_2) | \theta_1, \theta_2 \in [0, \pi]\}. \tag{8}$$

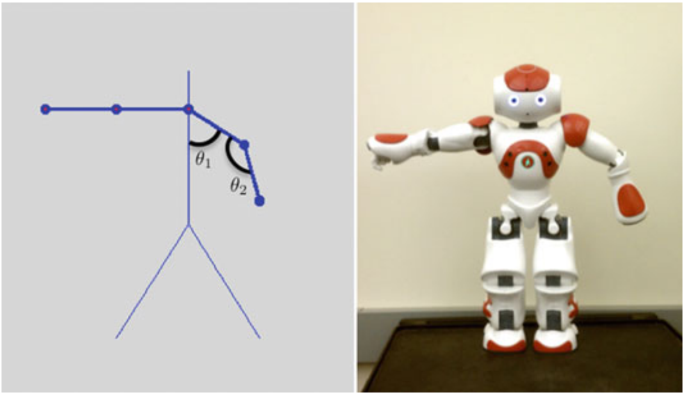


Fig. 8 The discrete states are interpreted as poses corresponding to two joint angles: shoulder and elbow as restricted to the body's coronal plane. On the *left* is the simulated view of a pose, and on the *right* is a corresponding pose on an actual robotic platform [11]

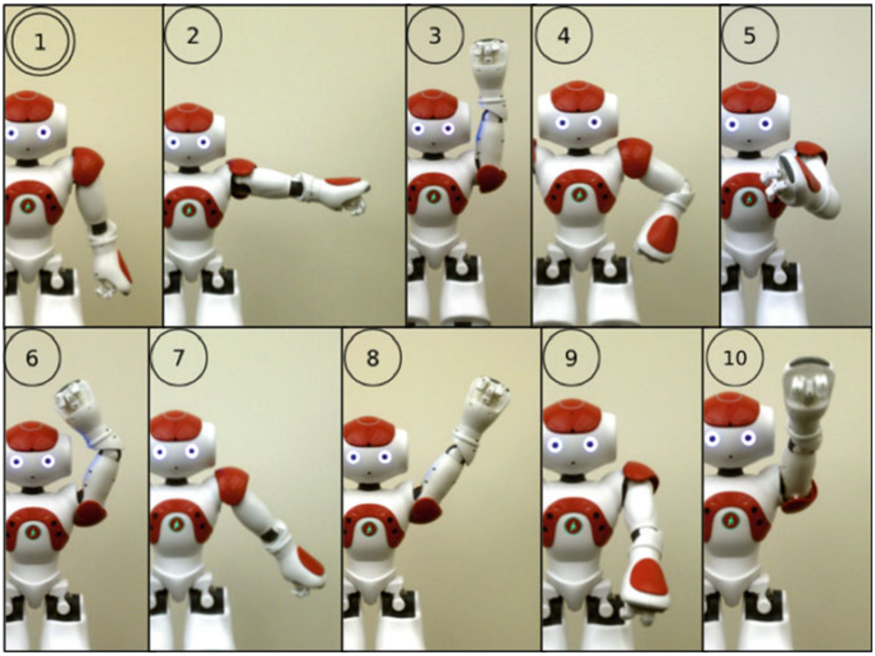


Fig. 9 An illustration of the ten arm poses chosen as those of importance and their corresponding discrete states which are used throughout this section—namely, in Figs. 10 and 11 [11]

The state transitions (events) correspond to movements, i.e., “put hand on hip” and “extend arm straight out.” That is, here, we enumerate primary motions that in this case are basically linear interpolations between their respective start and end poses. As in our prior work [11], X^{disco}_{arm}, E^{disco}_{arm}, f^{disco}_{arm}, and Γ^{disco}_{arm} are defined in Fig. 10. Moreover, the rest position is given by the state labeled “1.” As the system should always aim to return to that pose, the initial state is given by $x_0 = 1$, and the marked states are set to be $X_m = \{1\}$. The resulting marked language, i.e., the set of event strings that start at x_0 and end in X_m, produce feasible pose sequences recognizable as being in a style of disco dance. Likewise, X^{cheer}_{arm}, E^{cheer}_{arm}, f^{cheer}_{arm}, and Γ^{cheer}_{arm} are defined in Fig. 11. Note that $\mathcal{G}^{disco}_{arm_{1,2}}$ and $\mathcal{G}^{cheer}_{arm_{1,2}}$ have the same initial and marked states.

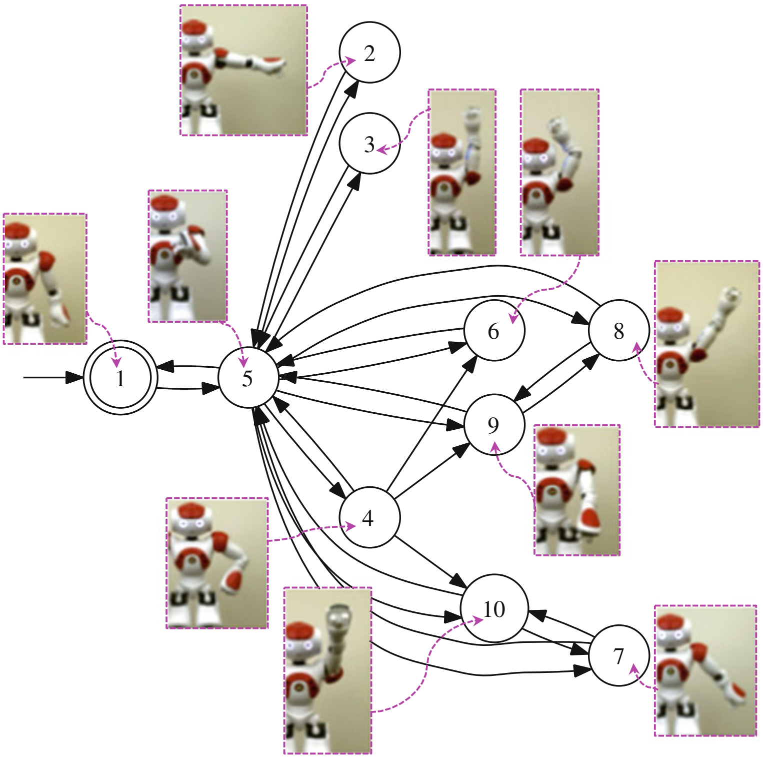


Fig. 10 A discrete event model of one arm performing the “disco” behavior. States correspond to poses defined by two joint angles: shoulder (as limited to the coronal plane) and elbow. Events are given by primary movements plus the empty event (or hold), which corresponds to undrawn self-loops [11] system and action space of the robot platform

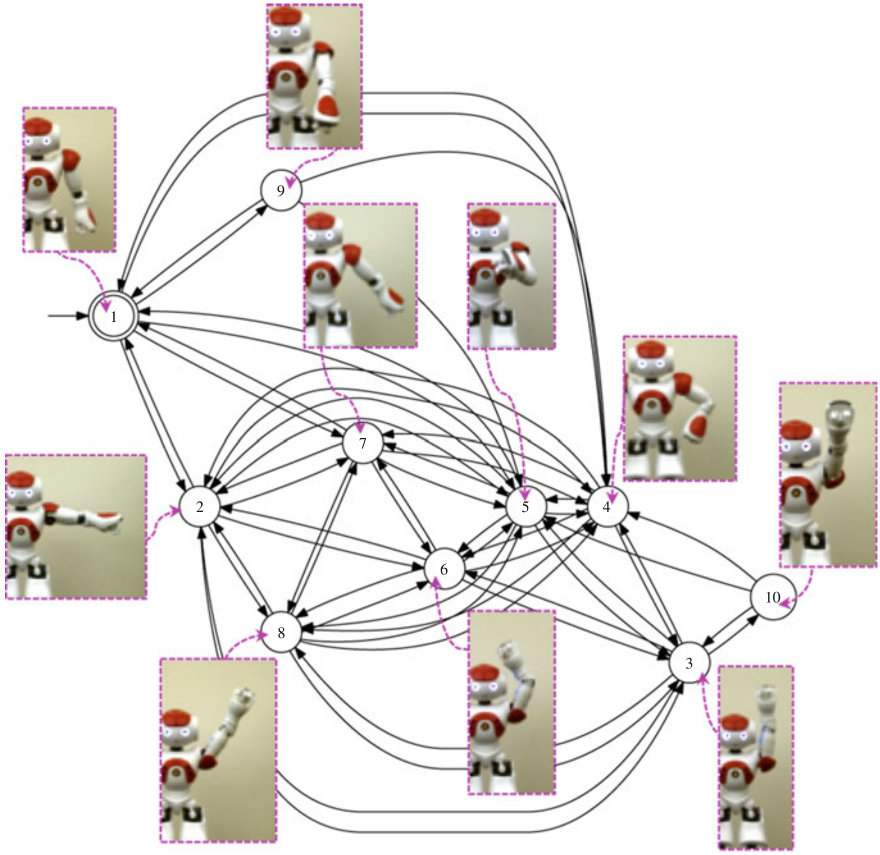


Fig. 11 A discrete event model of one arm performing the "cheer" behavior. States correspond to poses defined by two joint angles: shoulder and elbow. Events are given by primary movements plus the empty event (or hold), which corresponds to undrawn self-loops [11]

Finally, the framework presented here is able to scale the timing of these trajectories with the stylistic weights or knobs we outlined previously. Namely, we consider a 4-dimensional system with double integrator dynamics where $x = [\theta_1, \dot{\theta}_1, \theta_2, \dot{\theta}_2]^T$, $u = [u_{\theta_1}, u_{\theta_2}]^T$, and

$$\dot{x} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix} x + \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 0 \\ 0 & 1 \end{bmatrix} u \tag{9}$$

$$y = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} x. \tag{10}$$

Note that $y(t)$ describes the pose of the arm over time and that these dynamics simply represent the relationship between accelerations and velocities of joint angles rather than physically meaningful quantities. We then select the following weight matrices for Eq. (2):

$$\begin{cases} Q_{disco} = 0.1 \cdot I \\ R_{disco} = 0.1 \cdot I \\ P_{disco} = I \\ S_{disco} = 100 \cdot I \end{cases} \tag{11}$$

and

$$\begin{cases} Q_{cheer} = I \\ R_{cheer} = I \\ P_{cheer} = 10 \cdot I \\ S_{cheer} = 100 \cdot I \end{cases} \tag{12}$$

where I is the identity matrix. The nominal movement reference signal, r, which Eq. (2) encourages the system to track, is simply the linear interpolation between the desired end poses given by the motion sequence.

As in our prior work [11], the relative scaling of these matrices reflects the fact that we'd like the disco behavior to be loose but confident; hence, we employ a small weight on trajectory following (Q_{disco}) and a large one on end point matching (S_{disco}) to produce a trajectory that is, in Laban's terms, flexible and bound. We would also like this behavior to be more energetic. Thus, we employ a small weight on the input (R_{disco}) and change in state (P_{disco}) creating trajectories which are also strong and sudden. On the other hand, the cheer behavior is more rigid but still confident (which implies a larger Q_{cheer} and large S_{cheer}) and energetic but somewhat sustained (larger R_{cheer} and P_{cheer}). Thus, in the LBMS framework, according to our mapping, these trajectories are direct, bound, strong, and sustained. The system described in the next section will abstract away the role of these matrices entirely and simply rely on the more intuitive language of LBMS.

5 A System for Design-by-Humans of Heterogeneous Behaviors

In the system we outline now, this mapping process will be achieved through a more intuitive interface. Instead of painstakingly prescribing the structure of discrete event systems, users will select from a menu of motion options, which will be checked by the system for feasibility. And, rather than enumerating the entries in potentially large matrices, users will be able to adjust the Effort of actions taken by the robot through up and down arrow selectors on a web-based interface. The robotic platform we will control is a Rethink Robotics Baxter Research Robot pictured in Fig. 12.

Fig. 12 The Baxter robotic platform in the RAD Lab; manufactured by Rethink Robotics. The platform has: 3 onboard cameras (end of each arm and LCD display); electric parallel grippers, with downloadable CAD files for end effector customization; 360° Sonar specifically designed for human collision detection; 7 degrees of freedom (DOF) dual arm configuration; force sensing actuators, employing series elastic motors; Pedestal mounted with 4 locking casters for easy maneuverability; and is programmable in the Robot Operating System (ROS) environment

5.1 The Robot Operating System (ROS)

Robot Operating System (ROS) is an open source framework for robot software development providing operating system-like functionality. We will utilize it in our novel interface and provide a review of it here to demonstrate the type of use case our system replaces. ROS comes with client libraries for C++, Python, LISP and Java. What follows is a typical scenario for accessing a program written in Python using a set of commands provided by ROS. Figures 13, 14 and 15 and their captions detail the process step-by-step.

From this example, the level of computer expertise required to use ROS is clear; parameters are hidden within the bowels of low-level Python scripts and the Command Prompt is used to run these applications. In the next section, we propose a web-enabled system that provides a higher level of abstraction which removes the

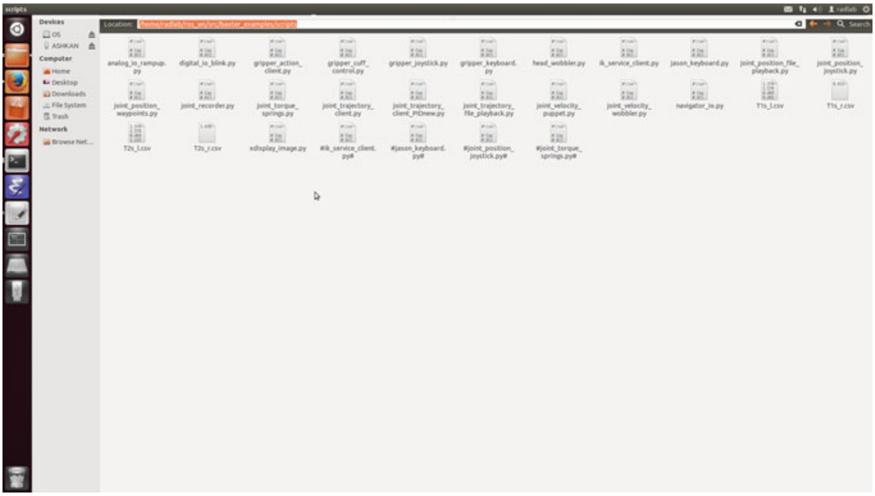


Fig. 13 Scripts and functions stored as files in a Linux-based machine. First, the user opens a terminal and navigates to the ROS workspace

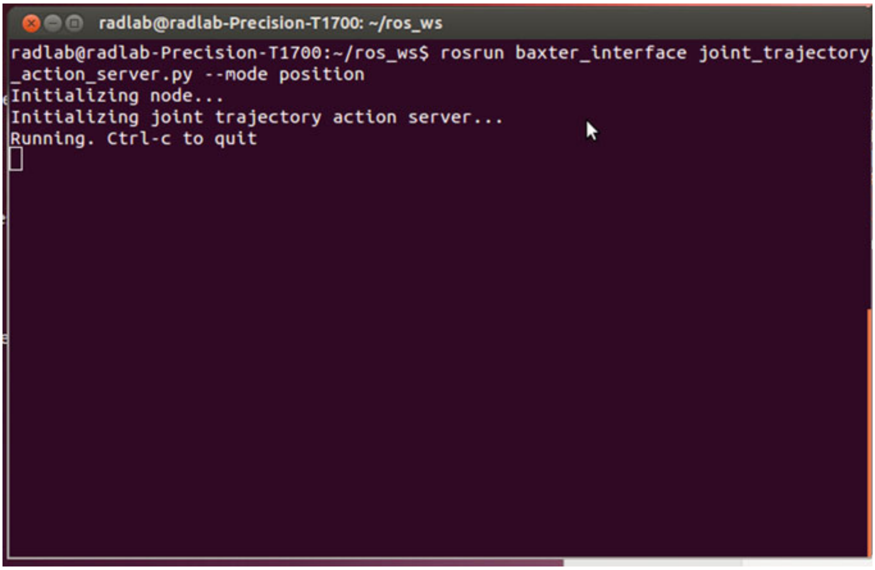


Fig. 14 Running the server program. Next, the user can run the server program using the commands provided by ROS. The user runs the desired program using the rosrun command which is one of the many commands provided by ROS. The server program waits for a client to connect

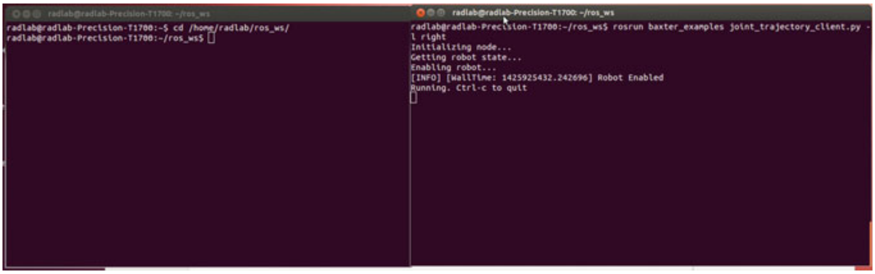


Fig. 15 Navigating to the ROS workspace and running the client. Finally, to execute the framework presented here, the user needs to run the client program that takes the generated trajectories by the MATLAB controller program and sends them to the server which passes them along to the robot. The programs use functionalities provided by ROS to communicate with each other and the robot

need for the user to directly interact with the ROS environment. The web application will take control of the interaction with ROS and the MATLAB controller program and let the user pick the parameters.

5.2 A Higher-Level of Abstraction, Our System

The user interface is a web application written in Java. The application provides a graphical user interface that allows the user to choose a set of pose transitions from a list that offers four different types of transitions, namely Flexion, Extension, Gathering and Scattering. These movements and their discrete event structure are shown in Fig. 16. The user can also manipulate the quality of movement by changing the control parameters of the controller. The four control parameters are associated with words describing specific aspects of LBMS Effort. A picture shows the feasible motions supported by the system in terms of a finite state machine where the transitions are annotated by the motif symbols. The user can pick any feasible series of pose transitions and tune the overall movement quality parameters to achieve a styled movement. Due to its high-level nature and focus on creating sequences of motion, we call this system the Robot Control/Choreography Center (RCC) (Fig. 17).

5.2.1 User Interface

A typical scenario for controlling the robot's movement from the user's point of view is as follows.

1. **Picking the desired motion transitions**. The user picks an unlimited number of motion transitions from the PickListMenu.
2. **Verifying the Movement**. Next the user clicks on the Verify Button to check the feasibility of the selected motions. If the selected motions are feasible the

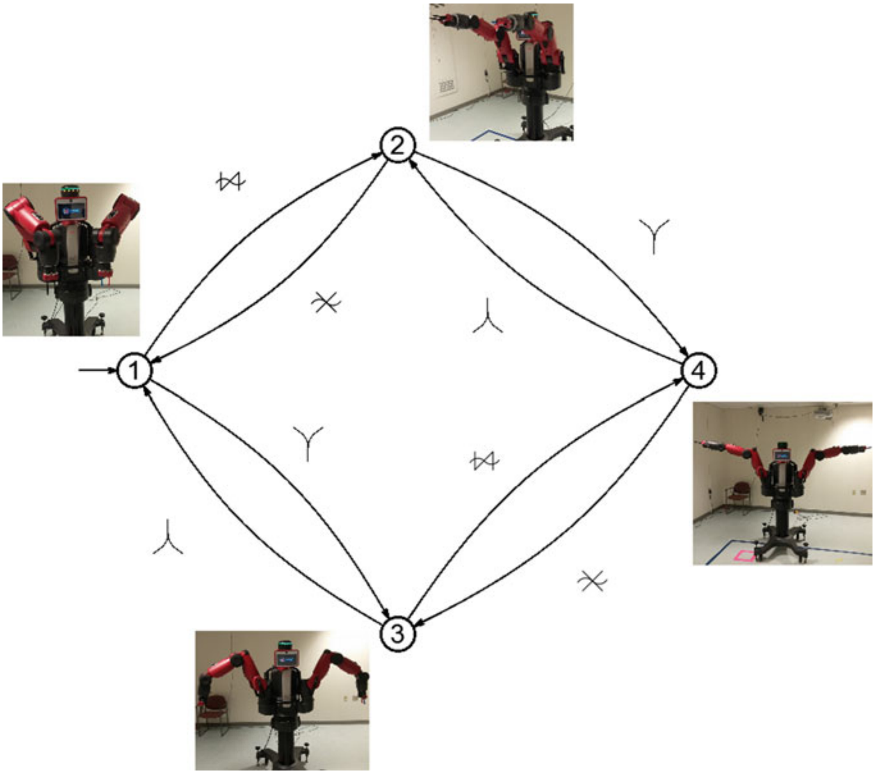


Fig. 16 The discrete event description of movement for currently implemented in the system. A general example that could be made more specific for particular scenarios and tasks. The events in the system are labeled with motif symbols for quick description of the type of primitive instantiated in between. Snapshots of the robot itself are shown next to the corresponding states

Submit button will be enabled and a message will show up on the upper right corner of the screen with the text "Feasible Movement." If the motions are not feasible, another message will show up listing the infeasible selected motions. In this case the user will need to remove the infeasible motions from the selected motions or add other motions to the list that make the overall movement feasible.

3. **Selecting movement quality parameters**. After successfully verifying the selected motions, the user can tune the movement quality parameters. The user selects the desired values for each parameter while being assisted by the system through an updating message after each spinner showing the quality hinted by the values selected for each parameter.
4. **Submit**. After selecting the desired movement quality the user clicks on Submit to send the commands and parameters to the controller. A message will show up on the upper right corner listing the selected motions and movement qualities.

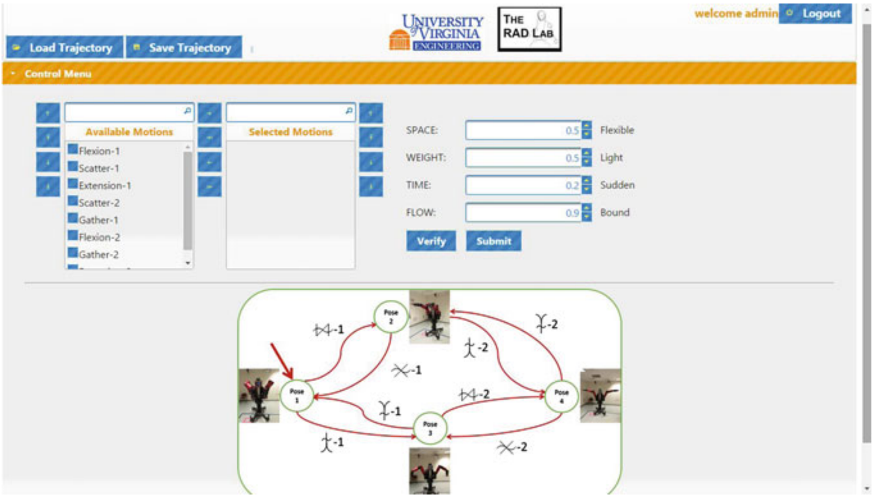


Fig. 17 Screenshot of the Robot Control/Choreography Center (RCC). At *bottom* is a view of the discrete event representation of movement; at *top* are the various controls, zoomed in Fig. 18. The interface requires no knowledge of command line prompts or low-level specific platform settings

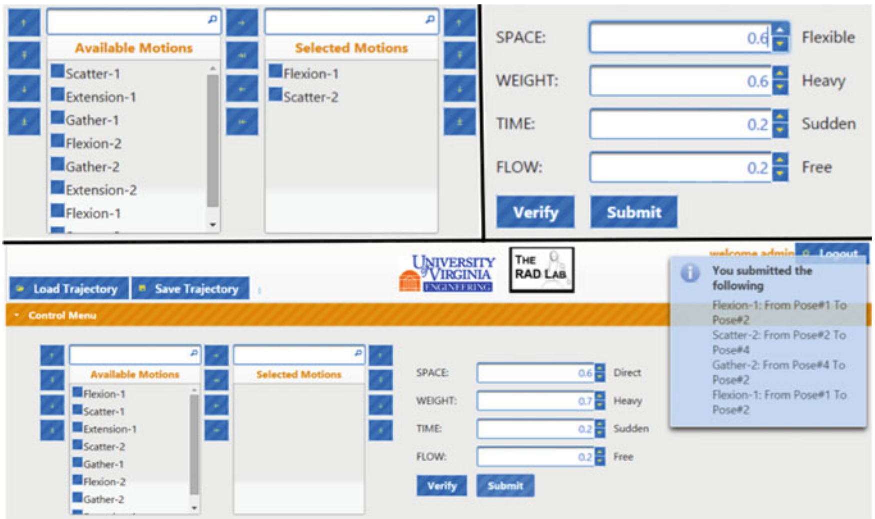


Fig. 18 *Upper left* PickListMenu (for Step 1). *Upper right* Quality parameters (for Step 3). *Lower* Overview, once the user has submitted movements (Step 4). The interface requires no knowledge of command line prompts or low-level specific platform settings

5. **Save Trajectory**. The user can save the current styled trajectory by clicking on the "Save Trajectory" button. The stored trajectories can be retrieved at any time by clicking on the "Load Trajectory."

The web application itself is developed with JavaServer Faces (JSF) which is a Java specification for building component-based user interfaces for web applications. The logic of the application is written in Java (JDK 7 Standard Edition). The application uses PrimeFaces library to create a more user friendly environment. PrimeFaces is a component suite open source User Interface component library for JavaServer Faces based applications. The application is deployed on Apache Tomcat 7, which is an open source software implementation of the Java Servlet and JavaServer Pages technologies and acts as the web server of the web application.

The RCC is comparable to the Aldebaran software suite *Choreographe* for the NAO platform. However, our system is at a higher-level of abstraction—making it easier to use but providing the user with a lower degree of control over the robot motion, relying more heavily on preprogrammed options than in the *Choreographe* suite. In our system, all the user sees is the web interface and the robot. The control part described in the next section is not seen by the user. In our system, the user makes the choice of what he/she wants the robot to do in the interface and then the robot will follow the instruction. The low-level controller ensures that the trajectory generated by the framework described in Sect. 3 is achieved for the particular platform; more details about this process are described next. Future iterations of this work will involve multiple platform-specific low-level controllers used extend the system to multiple robot platforms.

5.2.2 Low-Level Control

To generate the stylized motion trajectory for Baxter to execute, we need to define two things: the sequence of motions and the style of movements. They are picked by the user in the web interface. The sequence of motions tells Baxter where it should move. It is the transition between preselected poses and gives a draft of the trajectory. It is defined based on the objective of the movement. In our system, there are four states. State 1 is the pose that Baxter bends both of his arms in front of the body. State 2 is the pose that Baxter extends both arms so that the arms point to the front. State 3 is the pose that Baxter scatters the bended arms in State 1 to both sides of left and right. State 4 is the pose that Baxter extends both arms and scatters to both sides. Each of the arms in Baxter has seven degree-of-freedom (DOF). We record the joint positions for each state and hard-code them as the shape library in the program, leveraging on-board capability of the Baxter platform for intuitive primitive instantiation.

The style parameters described previously tell Baxter precisely how it should move, which means it defines the shape of the trajectory between the starting state and the ending state. It can be described by the weights in the cost function, which corresponds to the Effort system laid out by Rudolf Laban. The weights are denoted as (Q, R, P, S), which describes the cost of following the reference trajectory, the cost on control input, the cost on changes in states and the cost on terminal poses separately. Different weights will result in varying trajectories between the poses.

With different combination of sequence of motions and weights, we can generate significantly different styles of robot motion for a variety of objectives, such as passing an object, making a welcoming pose and so on. We used MATLAB to solve the optimal control problem to get the optimal joint angles in arms of Baxter to follow the reference trajectory. Those optimal joint angles were saved in a file and loaded to the interface of Baxter. After we generated the reference trajectory information for Baxter to follow, we loaded the position information and sent the corresponding commands to the controller to create a smooth trajectory. We used Python SDK packages to control Baxter.

A proportional-integral-derivative (PID) controller was used for the motors. The current positions of joints were collected as the negative feedback shown in Fig. 19, and we calculated the differences between the measured positions and the desired positions. The controller tries to minimize the difference between its actual trajectory and the one prescribed by the high-level controller. Thus, each joint in Baxter tracks its independent reference trajectory.

We used joint trajectory action server which comes with Baxters SDK in the package of `joint_trajectory_action` to command Baxters arms to go through the desired points with a PID controller. The desired position of each joint was generated by MATLAB controller, loaded in the joint trajectory client at a series of specific time and sent to the joint trajectory action server. The PID gains can be configured to get a satisfactory dynamic response with an overshoot less than 10 % and steady state error less than 5 %. In our system, to make Baxter move between the desired states with the given performance and the joint angles track the commanded trajectories, the PID gains can be chosen as (100, 500, 10).

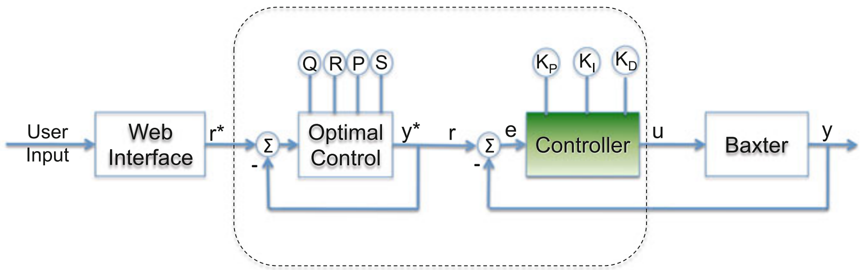


Fig. 19 Architecture of the low-level controller utilized in this framework

6 Application in Advanced Manufacturing, Toward Personal Robotics

We know that we can create diverse sets of behaviors by sequencing and weighting motion primitives. Let us now take a look at the abstraction of sequencing behaviors and the implication of interaction on a diverse, or heterogeneous, team of agents. This is important for formally modeling the robot's behavior and to show dependencies of a team with multiple robots, humans, and machines. We use these 'abstract machines' to describe our sequence of behaviors by allowing the motion primitives to be represented by arrows and the states to be depicted by circles.

In an collaborative environment where a robot's behavior may be coordinating with other agents, we can describe the various members of the team with this same framework. When adding team members in a manufacturing setting, the set of such agents might include but is not limited to: Humans, Robots, Fixtures, and Machines. We can then construct interaction graphs in order to model dependencies, or how each entity interacts with other entities in a team. This is particularly important to aid in understanding the interaction a human agent has with a robot agent, something that manufacturers care a great deal about. They can use the interaction, or dependency, seen on the graph to understand process implications with regards to safety and other concerns. The dependencies can also be assigned weights, upper bounds, and lower bounds to allow for further analysis. The complete set of defined agents, their behaviors, and interaction graphs comprises a behavioral design for such a mixed a team. The concept is to allow for methodically creating multiple designs and comparing them alongside each other with respect to dependency and other selected objectives.

We will now describe one specific method used for defining sequenced robot behaviors and structuring each agent into a team. The five stages of the method are outlined as follows.

1. **Introduce agents**. List the separate entities that will take part in the team of interest. These entities need to correspond to the members of the dance team, work-cell, or whichever group you wish to study. If you do not yet have a team in mind, it is possible to model an individual entity process before an entire team. It is good to keep in mind that creating new or disposing of old entities may be necessary as multiple designs are eventually produced.
2. **Construct discrete event description with respect to each agent, to define a library of motion primitives and states**. In order to build process automata, you will need to define all the motion primitives and states necessary to complete the action. This requires the formulation of an overall team goal that could be a range of activities from a ballet performance to a manufacturing task. Once a team goal is specified, create the processes for each entity, detailing the actions they perform along with states achieved. Examples of an action for a robot entity might be picking up a part or moving the left arm to a new position. Actions will be represented as arrows, with the tail of the arrow leaving state x_i and moving

to state x_{i+1}. States are portrayed as circles, with a double circle indicating the final or accepting state.

3. **Construct interaction graphs for the complete team to show agent dependencies**. For the selected team write down all of the previously defined entities from Stage 1 as vertices in a graph $\mathscr{G}(V, E)$. The vertices should be arranged by placing an entity that interacts with different agents adjacent to each other. Draw a directed edge originating from the entity that is performing an action upon another entity. This is best completed by following the actions and states for each process automata created in Stage 2, and adding directed edges whenever an interaction occurs.
4. **Create multiple designs of unique processes and interactions**. After Stages 1 through 3 are complete for the initial design, it will be possible to move onto creating a new design. This new design can allow for either changing the process with existing entities or, a more interesting case, introducing new agents such as another robot or human to the team abstraction. There is no defined stopping criterion suggested for knowing the minimum or maximum number of designs. We must rely on the subject matter experts to create as many designs as necessary for a complete ad hoc analysis.
5. **Compare the iterations of all created designs against objectives**. Our method, while possibly generalizable to more objectives, is well suited to understanding a team's interaction complexity. This is particularly useful when implementing collaborative robots that will exist alongside humans. A factory safety specialist, for example, would be able to infer which design has a greater human-robot dependency.

Thus, this very basic framework is flexible and high-level enough to capture ***the many***, ***varied*** contexts where we may want to employ robots—be it on the manufacturing floor, the proscenium stage, or in our own homes. Furthermore, it ***lowers the barrier to entry*** for crafting such behaviors, allowing a much wider class of users such as line workers in a factory, dancers in a studio, and individuals in their own houses. Such tools are essential to realizing the potential of robots in society and require a combination of embodied movement theory and tools from robotics and control.

7 Conclusion

This chapter has reviewed prior general concepts in robotics and somatic theory. Further, we have presented our prior work that provides a foundation for one way of thinking about how these concepts are naturally related. Namely, we have showed how events in finite state machines can be modifiable movement primitives described by Motif symbols from LBMS. This way of thinking provides a natural abstraction for the development of a novel system—presented here—that can facilitate the control of and programming of robots by nontechnical persons.

The RCC is not only an important marriage of academic ideas but is also an important practical tool for collaboration between the disciplines of dance and robotics. It provides an initial tool to allow dancers to be the designers of machine movement and a natural translation between the distinct languages of LBMS and robotics. It is through such tools that experts in movement will be able to influence and improve the way that robots impact our lives.

References

1. R. Arkin, *Behavior-Based Robotics* (The MIT Press, Cambridge, 1998)
2. K. Studd, L. Cox, *Everybody is a Body* (Dog Ear Publishing, Indianapolis, 2013)
3. V. Maletic, *Body, Space, Expression* (Walter de Gruyter & Co., Berlin, 1987)
4. A. Hutchinson, A. Guest, *Labanotation, or, Kinetography Laban: The System of Analyzing and Recording Movement* (Dance Books, 1996)
5. I. Bartenieff, D. Lewis, *Body Movement: Coping with the Environment* (Routledge, 1980)
6. P. Hackney, *Making Connections: Total Body Integration Through Bartenieff Fundamentals* (Routledge, 1998)
7. J. Newlove, J. Dalby, *Laban for All* (Nick Hern Books, London, 2004)
8. A. LaViers, M. Egerstedt, The ballet automaton: a formal model for human motion, in *Proceedings of the 2011 American Control Conference* (2011)
9. A. LaViers, Y. Chen, C. Belta, M. Egerstedt, Automatic generation of balletic motions, in *ACM/IEEE Second International Conference on Cyber-Physical Systems* (2011), pp. 13–21
10. A. LaViers, Y. Chen, C. Belta, M. Egerstedt Automatic sequencing of ballet poses. Robot. Autom. Mag. IEEE **18**(3), 87–95 (2011)
11. A. LaViers, M. Egerstedt, Style based robotic motion, in *Proceedings of the 2012 American Control Conference* (2012)
12. C.G. Atkeson, S. Schaal, Robot learning from demonstration, in *Proceedings of the 14th International Conference on Machine Learning, ICML'97* (1997) pp. 12–20
13. B. Akgun, K. Subramanian, Robot learning from demonstration: kinesthetic teaching vs. teleoperation (2011)
14. G. Warren, S. Cook, *Classical Ballet Technique* (University of South Florida Press, Gainesville, FL, 1989)
15. A. Hutchinson-Guest, *An Introduction to Motif Notation* (Language of Dance Centre, London, 2007)
16. A. Hutchinson-Guest, *Motif at a Glance!* (Language of Dance Centre, London, 2000)
17. T. Nakata, T. Mori, T. Sato, Analysis of impression of robot bodily expression. J. Robot. Mechatron. **14**(1), 27–36 (2002)
18. M. Sharma, D. Hildebrandt, G. Newman, J. E. Young, R. Eskicioglu, Communicating affect via flight path exploring use of the laban effort system for designing affective locomotion paths, in *2013 8th ACM/IEEE International Conference on* IEEE *Human-Robot Interaction (HRI)* (2013), pp. 293–300
19. H. Knight, R. Simmons, Expressive motion with x, y and theta: Laban effort features for mobile robots, in *The 23rd IEEE International Symposium on Robot and Human Interactive Communication, 2014 RO-MAN* (IEEE, 2014), pp. 267–273
20. A. LaViers, M. Egerstedt, Style-based abstractions for human motion classification, in *ACM/IEEE 5th International Conference on Cyber-Physical Systems* (2014)
21. Y. Sheng, A. LaViers, Style based human movement segmentation, in *IEEE International Conference on Systems Man and Cybernetics* (2014)

Annotating Everyday Grasps in Action

Jia Liu, Fangxiaoyu Feng, Yuzuko C. Nakamura and Nancy S. Pollard

Abstract Grasping has been well studied in the robotics and human subjects literature, and numerous taxonomies have been developed to capture the range of grasps employed in work settings or everyday life. But how completely do these taxonomies capture grasping actions that we see every day? In a study to classify all actions during a typical day, we found that single entries in an existing grasp taxonomy were insufficient, apparently capturing not one grasp, but many. When we investigated, we found that these seemingly different grasps could be distinguished by features related to the grasp in action, such as the intended motion, force, and stiffness. In collaboration with our subjects, we developed an annotation scheme for untrained annotators to use, which captured the differences we observed between grasping actions. This chapter describes our annotation scheme. We discuss parallels to and differences from Laban Movement Analysis, which has been long developed to capture motion and action, but does not focus on grasping. We also discuss parallels to impedance or operational space control, with the goal of moving from annotations to actionable robot control.

1 Introduction

Grasping is an essential part of people's daily lives and is critical for creating robots that can interact with and make changes to their environment. Grasping has been the focus of numerous human studies [1], and a large body of robotics research has worked within a grasp-move-ungrasp paradigm. Within these studies, one area of focus has been hand shape and the locations of contact between hand and object, which determine how the hand and object can interact [2].

J. Liu (✉) · F. Feng · Y.C. Nakamura · N.S. Pollard
School of Computer Science, Carnegie Mellon University, 5000 Forbes Ave, Pittsburgh, PA 15213, USA
e-mail: jialikeswinter@gmail.com

J.-P. Laumond and N. Abe (eds.), *Dance Notations and Robot Motion*,
Springer Tracts in Advanced Robotics 111,
DOI 10.1007/978-3-319-25739-6_12

A number of taxonomies with hand shape and object contact as central elements have been developed to classify grasps [2–5]. These taxonomies have been widely used in robotics, for applications including grasp recognition [6, 7], robot hand design and evaluation [8], programming by demonstration [9], and interaction with grasp sensitive objects [10]. These taxonomies also allow researchers to communicate grasp differences, distinguishing power from precision grasps, tripod versus pinch, spherical versus cylindrical, etc.

With the goal of moving toward robots that can more dexterously manipulate everyday objects in human environments, we ask to what extent our existing grasp taxonomies capture the actions we do in everyday life. Two subjects attempted to capture all actions accomplished during a typical day, with a focus on critical humanoid robot capabilities such as home care and manipulation in unstructured environments such as a home or workplace. For each observed grasp or manipulation action, our subjects attempted to classify it using the Comprehensive Grasp Taxonomy of Feix et al. [5]. In all, 179 distinct grasping actions were captured and classified.

We found that although many grasping actions could be classified in the existing taxonomies, there were important differences between grasps that the taxonomy did not consider. To capture these differences, we propose an extended set of annotations capturing aspects of force, motion, and stiffness. Table 12 shows an example. Our goal was to communicate motion, force, and stiffness information as precisely as possible while still allowing individuals with light training to understand and classify grasps or communicate differences to a robot.

We found 40 grasp types which could not be well captured by existing taxonomies, including actions of pushing, grasping while pressing a button or lever, and grasping with extension (inside-out) forces. We believe our database is an improvement on our prior work, because we characterize human grasps by taking into account forces and motion exerted after a grasp is achieved. These added properties have intriguing similarities to aspects of dance notation such as Laban Movement Analysis [11]. They also may tie into existing impedance [12] and operational space controllers [13] used in robotics.

We report our complete process and findings below. The complete classification can be viewed in our online database [14]. A short version of this paper has previously appeared elsewhere [15].

2 Related Work

Perhaps the earliest well known grasp taxonomies are those of Schlesinger [16] and Napier [17], which led the way in discriminating major hand shapes and grasp functions. Grasp taxonomies have been developed for tasks of everyday living, including those of Kapandji [18], Edwards et al. [4] and Kamakura et al. [2]. Kamakura and colleagues, for example, classified static prehensile patterns of normal hands into 14 patterns under 4 categories (power grip, intermediate grip, precision grip and grip involving no thumb). They illustrated detailed contact areas on the hand for each grasp and analyzed for which objects the grasp may be used.

Perhaps the most widely cited taxonomy in robotics is that of Cutkosky [3], which includes 16 grasp types observed in skilled machining tasks. The Cutkosky taxonomy consists of a hierarchical tree of grasps, with categories classified under power and precision. Moving from left to right in the tree, the grasps become less powerful and the grasped objects become smaller. Zheng et al. [19] used this taxonomy to capture the daily activities of a skilled machinist and a house maid, giving for the first time a count of how frequently different grasps are used. The intent of our study is similar. However, we consider a broader variety of actions beyond static grasps and make special note of differences observed in grasps that have the same entries within the grasp taxonomy.

Feix et al. [5] recently developed a comprehensive taxonomy of grasps that brings together previous research with their own observations. They propose a definition of a grasp as follows: “A grasp is every static hand posture with which an object can be held securely with one hand,” and identify 33 grasp types that are distinct from one another and fit this definition. Because it was developed with the goal of being inclusive, we selected this taxonomy as a starting place in our experiments. However, the grasp definition of Feix et al. [5] does exclude a variety of movements, bimanual tasks, gravity dependent grasps, and flat hand grasps that we found important, and we include these additional types in our own taxonomy.

A number of taxonomies have been developed to express manipulation actions as well. Chang and Pollard [20] classify manipulations prior to grasping, with a focus on how the object is adjusted, considering both rigid transformation and non-rigid reconfigurations. Worgotter et al. [21] discuss how manipulation actions are structured in space and time. Focusing on actions of bringing together and breaking apart, they identify 30 fundamental manipulations that allow sequences of activities to be encoded. Elliott and Connolly [22] classify coordinated motions of the hand that are used to manipulate objects, identifying three classes of intrinsic movements: simple synergies such as squeeze, reciprocal synergies such as roll, and sequential patterns such as a rotary stepping motion of the fingers to change contact positions on the object. Bullock et al. [23] encode manipulation instances at a more abstract level, focusing on motion of the hand and relative motion of the hand and object at contact, with the goal of creating a classification scheme that does not assume a specific hand design. We adopt a structure similar to theirs for expressing intended motion of grasped object, but incorporate it as extra information within the context of a more conventional grasp taxonomy.

Torigoe [24] investigated manipulation in 74 species of great apes, identifying over 500 different body part manipulation acts, 300 of which are related to hand manipulation, including drape, flip, pick up, pull, push or press, roll, rotate, throw, untwist and so on. We find that a similar approach of classifying manipulation actions using action verbs is useful for distinguishing between different force intentions for grasps having the same grasp taxonomy label and adopt it as extra information in our taxonomy.

3 Methods

We compiled a task list from various sources for our study. First, we studied previous literature that measured self-care and mobility skills for patient rehabilitation [25–28]. The measured skills listed in these papers such as dressing, eating, and grooming cover typical and important tasks humans need to do, even for those who are disabled. Our initial list of actions was a union of the tasks mentioned in those papers. In work such as Choi et al. [29], tasks were ranked by importance, and tasks like buttoning, putting on socks, and personal hygiene were discarded because they received a low ranking and are difficult for a robot to accomplish. However, we also included these less important tasks in our list, with the goal of having a more inclusive study.

We next observed two college students' life from the time they woke up until the time they went to bed. We categorized all the hand gestures and motions that the person would use into hundreds of tasks. However, we felt this was insufficient since there are many skilled gestures (e.g. of tradespeople) that are not found in everyday life, and that the task list was biased toward the office settings of the subjects. Therefore, we expanded our task list to include specific tasks that people from various careers would accomplish in their workplace.

Next, we further separated the compound tasks into small task-components and movement elements, such as in Kopp et al. [25]. For example, wearing a T-shirt was broken down into three basic tasks: (1) arms in T-shirt sleeves, (2) grab the neck hole and move head through neck hole, (3) pull down and straighten shirt. We collapsed similar gestures together and classified these movements into an existing 33-grasp database [5]. When we encountered hand gestures that were not in the basic database, we added them to the database.

Our final database contains 73 database categories, of which 50 are grasp types, 4 are press types, 10 are grasp and press type, 2 are extend types and 7 are other hand types. We also illustrate where each movement may be used in daily life with corresponding pictures [14].

4 Database Annotations

Figure 1 shows the classification we have developed in order to distinguish the different actions we have observed. The focus of previous grasp taxonomies has often been on hand shape (highlighted in purple). With our observations, however, we annotated grasps with four features: (1) hand shape, (2) force type, (3) direction, and (4) flow. The object related property is another factor that influences the hand shape and motion, but these relationships are not made explicit in our database. In contrast to traditional grasp taxonomy research, our research focuses on grasps within the context of the action that is intended. The rationale behind this focus came about when we mapped the grasping actions we encountered onto an existing

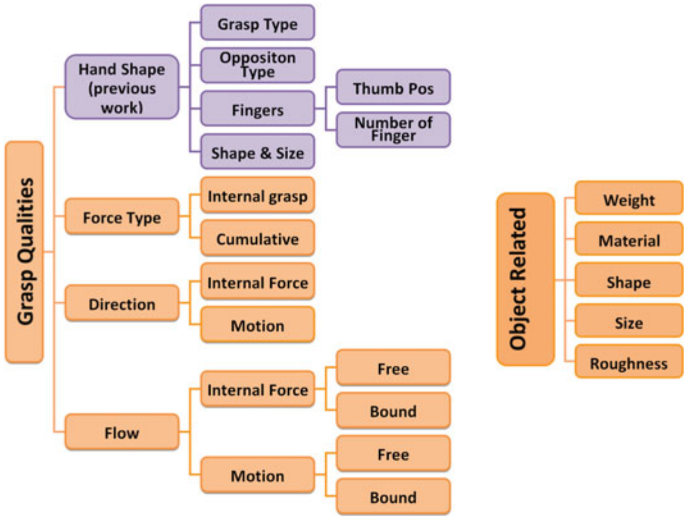


Fig. 1 Simple classification of the database

grasp taxonomy [5] and realized that actions belonging to one grasp type within the taxonomy often involved very different motion, force, or flow.

4.1 Hand Shape

Our classification of hand shape comes directly from Feix et al. [5], combined with ideas from Napier [30]. For hand shape, we consider: grasp type, opposition type, thumb position, involvement of specific fingers, and prototypical object shape and size.

Grasp type can be power grip, precision grip, or intermediate. A power grip is typically applied by partly flexed fingers with the palm providing countering pressure, while a precision grip is more of a pinching of the object between fingers, which allows freedom to sense or create internal motions of the object within the hand.

Opposition type refers to which part of the hand is mostly used, including palm (red in Fig. 2), pad (green), side (blue), and back (Fig. 3).

Thumb position is classified as abduction (ABD), adduction (ADD), extension (EXT), or flexion (FLX) (Fig. 4). It is also important to indicate specific fingers (2: index finger, 3: middle finger, 4: fourth finger, 5: little finger) involved in each gesture.

Finally, we express shape (spherical, cylindrical, disk-like, etc.) and size (large, medium, small) of the object being held [30].

Fig. 2 Palm, pad, side

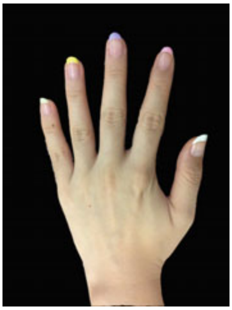


Fig. 3 Back

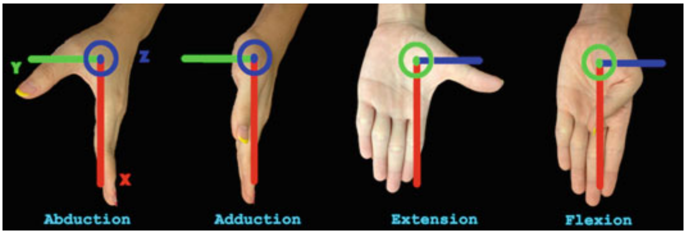


Fig. 4 Local coordinates and thumb positions of the left hand

4.2 Force Type

There are many different ways in which forces can be distinguished or described: axis direction, magnitude of the force, location of force exertion, and so on. However, we found that describing forces using verbs from the English language made it easier for our subjects to annotate grasping actions and provided a clearer description to other people than the alternatives we investigated. We use 20 verbs to describe the forces observed in our study (Table 1).

Table 1 Force type definitions and frequency

Force type	Definition	Frequency
Break off	Remove a part of an object	3
Extend	Apply outward forces from within the object	3
Grab	Hold or secure without opposing gravity	32
Hold	Grasp object in a way that resists gravity	41
Lever	Pivot one end of an object around a fixed end	4
Lift	Apply upward force greater than gravity	7
Place	Put something in a specified position	1
Press	Exert force in a direction away from the shoulder	31
Pull	Exert force in a direction towards the shoulder	18
Punch	Press or push with a short, quick movement	1
Put in	Insert one object into another	4
Roll	Cause rotation without prehension	3
Rub/stroke	Move back and forth while pressing	9
Scratch	Rub with something sharp or rough (with the hand directly or a tool)	2
Squeeze	Apply compressive force around object greater than needed to hold object	4
Swing	Move with a smooth, curving motion like hand waving or arm swinging	6
Take out	Remove one object from another	2
Throw	Propel an object through the air	3
Turn	Flip or rifle through pages	1
Twist	Cause rotation with prehension	13

Table 2 Internal force examples

Example		
Force type	Squeeze	Hold
Annotation	Squeeze toothpaste	Hold a pan

Although we don't make a distinction in our database, it's interesting to note that these force words imply (1) an internal grasp force exerted by the hand (e.g. *squeeze*, Table 2), or (2) a cumulative/external force exerted by the wrist or whole arm (e.g. *throw*, Table 3), or (3) both (e.g., *grab* a door handle and *press* to open, Table 3).

Table 3 Cumulative force examples

Example		
Force type	Throw	Grab and press
Annotation	Shoot a basket ball	Press down a door handle

Table 4 Hold and grab examples

Example		
Force type	Grab	Hold
Annotation	Grab the ladder	Hold a laundry detergent

In our database, both force and motion are important. For this reason, *grab* and *hold* are not the same, even though they feature the same motion (i.e. no motion). We define *grab* as touching or securing an object that is resting on a surface. We define *hold* with a gravitational factor, where the hand/arm is applying an upward force to counteract gravity (Table 4).

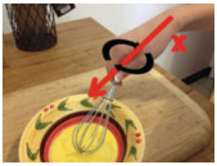
4.3 Direction

In order to specify the direction of a force or motion, we need to specify the direction subspace and the coordinate frame as shown in Table 5. The direction subspace describes a subset of the six-dimensional space within which the motion is occurring. Examples of direction subspaces that we use include: (1) along a linear axis, (2) rotation around an axis, (3) movement within a plane, or (4) inwards/outwards (towards or away from the center of an object). We note that the motion direction can be very different from the force direction. For example, when we zip a zipper, the internal force direction of the hand is *inwards* for the zipper (i.e. grab the zipper tightly), but the direction of motion is *along* the zipper. Similarly, the internal force direction is *inwards* to hold the egg beater but the direction of motion is around the x-axis (Table 6). We use the notation x(45)y to describe movements along an axis that is halfway between the x- and y-axes (e.g., Table 12, second row).

Table 5 Direction examples

Property	Possible values	Example
Direction subspace	Along x/y/z axis or combination	Table 6 1
	Rotate around x/y/z axis	Table 6 2
	Plane xy/xz/yz	Table 6 3
Coordinate frame	Hand	Table 7 1
	Global	Table 7 2
	Object	Table 7 3

Table 6 Axes examples

Example			
Motion axes	Along x/−x (object)	Around x axis (hand)	Along xz plane (hand)
Force axes	Inward, hold zipper	Inward, hold egg beater	Against the surface
Annotation	Zip a zipper	Beat with egg beater	Move a mouse

Directions that are less constrained or more difficult to describe are captured in freeform text (e.g., “a cone about the x-axis” or “various”).

Most of the time, we use the local coordinates of the hand to describe the direction of movement. However, we also sometimes use global coordinates of the world or local coordinates of the object, depending on its degree of usefulness.

Hand coordinates: The local coordinates of the hand are defined as follows: The direction of the four fingers is defined as the x-axis. The y-axis is defined as coming out of the palm in the ventral/palmar direction. The z-axis is defined as the thumb pointing away from the little finger for both hands (Figs. 4 and 5). This results in using either the left hand rule for left hand or right hand rule for right hand to compute the z-axis. This unorthodox use of coordinate frames results in symmetrical descriptions of movements and grasps using the two hands. Local

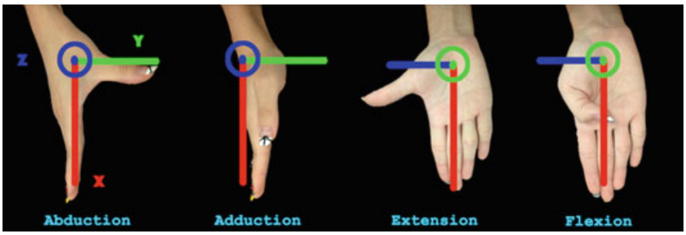


Fig. 5 Local coordinates and thumb positions of the right hand

Table 7 Coordinate frame examples

Example			
Coordinate frame	Hand	Global	Object
Motion axes	Along x/−x	Along z/−z	Along x/−x
Annotation	Rub hands	Dribble basketball	Measure with a tape measure

coordinates of the hand are mostly used when the motion is along one of the hand coordinate axes. For example, Table 7, first column, shows rubbing the hands along the local x-axis.

Global coordinates: Global coordinates of the world are used when the motion is along the direction of gravity or within a coordinate system that could be fixed to our local environment. For example, when we dribble a basketball, we maneuver the ball within a coordinate frame fixed to the world, not the hand or the ball (Table 7, second column). The direction of gravity is defined as the global z-axis.

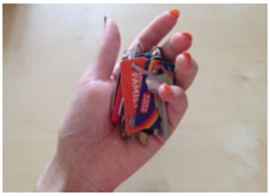
Object coordinates: Finally, occasionally the local coordinates of the object must be used since, in some motions, the object shape decides the direction of motion. If the object is a long stick or string type, we define the direction along the stick to be the x-axis. If the object is rectangular in shape, we define the direction along the long side to be the x-axis and the direction along the short side as the z-axis. For example, when we pull out measuring tape, the motion direction is along the tape's long dimension: the x-axis (Table 7, third column).

Many motions or forces can be described naturally in multiple coordinate frames. For example, plugging in a charger could be expressed in the coordinate frame of the charger, the wall, or the hand. We asked our subjects to make the annotations that were most intuitive for them. The important point is that all three coordinate frames are useful, as different actions may focus on different frames of reference.

4.4 Flow

The effort factor we use here is flow. Flow comes from the Laban Effort/Shape notation [31]. It refers to "attitude toward bodily tension and control" and can be *free*, *bound* and *half-bound*. Free refers to the moving direction of the gesture being very casual, while bound refers to the action being very stiff or tightly controlled.

Table 8 Flow factor examples

Example		
Flow	Bound motion/bound force	Free motion/half bound force
Annotation	Stick a key in the door lock	Hold keys

The half bound annotation is used when the action is bound along one or more axes and free along the rest. For example, in Table 14, the flow of motion in dragging toilet paper is half-bound because in the plane that is perpendicular to the axis of the toilet paper, the motion is still free. Our informal observation is that most of the time we specify an action as being free or bound depending on whether the action includes a goal location. For example, if we try to plug in a charger into a wall or stick a key into a lock, the motion is bound, but if we just throw the key for fun, the action is entirely free (Table 8).

4.5 Object Related Factors

Most grasps depend on the object our hands manipulate, thus object related factors are also important features for describing hand gestures.

From our observations, weight is an important factor since it affects both internal and cumulative force applied on the object. A simple example is when we hold an empty box or a full box. If the box is empty, we tend to grab the top piece of the box, but if the box is heavy, we would hold from the bottom and lift it up (Table 9).

The material of the object also strongly affects grasping strategy. For example, grabbing highly deformable material requires continuous adjustment of grasp shape

Table 9 Weight of object examples

Example		
Object weight	Light	Heavy
Annotation	Grab an empty box	Hold a heavy box

Table 10 Shape and size and roughness of object examples

Example			
Size	Thin	Thick	Thick
Roughness	Slippery	Rough	Slippery
Annotation	Grab a wire	Grab a rope	Grab exercise bar

as the object changes shape. Another example of the effect of material is that people will grab raw meat differently than paper.

The shape and size of the object affects hand shape. We usually pinch a thin wire but grab a thick string, see Table 10.

Finally, the friction coefficient of an object determines how hard we grab the object. The thick string in Table 10 is rougher then the exercise bar, which will affect the force needed to prevent slipping in both cases.

5 Results

Our main result is an annotated database of grasping actions observed in our study. The database contains 73 grasp types, including the 33 types enumerated in Feix et al. [5], along with 40 additional types. Each of these 73 types includes one or more annotated examples. Examples are annotated with force type, motion direction, force direction, and flow to more fully describe the grasp in action. Each of the 179 total examples differs from the others by at least one annotation.

One additional result listed here is a count of force types, which can be found in Table 1 (frequency column). In this table, we can see, for example, that *hold* (41), *grab* (32), *press* (31) and *pull* (18) make up the majority of tasks that we observed in our study.

The full database can be found on our website [14]. In this Chapter, we describe two of the 73 grasp type entries (Sects. 5.1 and 5.2) as well as listing some of the new grasp types (Sect. 5.3).

5.1 Large Diameter Cylinder

The first grasp type we examine is the large diameter cylinder grasp. In a large-diameter grasp (Table 11, Left), the hand shape is appropriate for a larger-diameter cylinder-shaped object, and all five fingers are used. The opposition type is palm. The thumb is abducted.

Table 11 Large diameter and lateral grasp

Name	Large diameter	Lateral
Picture		
Type	Power	Intermediate
Opp. type	Palm	Side
Thumb pos	Abd	Add
VF2	2–5	2
Shape	Cylinder/cuboid	Card piece
Size	Large diameter	Thin

Our entire database entry for this grasp is shown in Table 12, and we see that this single entry in the grasp taxonomy contains a variety of different examples. Force types are varied, including *hold*, *grab*, *squeeze*, *press*, and *twist*. Even with the same force type, other annotations can differ. For example, as shown in Table 12 (top), the action of *drink water* involves motion around the y-axis, while holding a bottle does not involve any motion. The flow can vary even within the same task. As shown in Table 12 (bottom), the motion of squeezing a towel is free, but the force is bound.

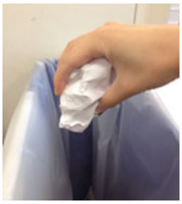
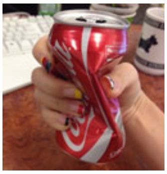
5.2 Lateral

The second grasp type we review is the lateral grasp. As shown in Table 11, Right, in the lateral grasp, the hand shape is more suitable for a thin card-shaped object, which is pinched between the thumb and index finger. The opposition type is side, and the pad of the thumb is used. The thumb is adducted.

For some very similar tasks, the direction and flow can be different. As shown in Table 13 first row, the flow of motion in putting on gloves and dragging toilet paper are different. Putting on gloves is bound since the direction of motion is set along the arm. But dragging toilet paper is half-bound.

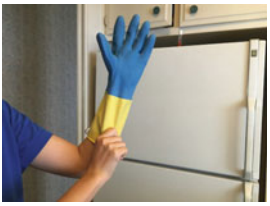
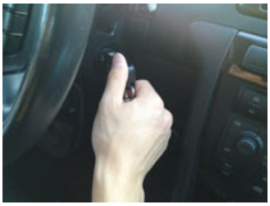
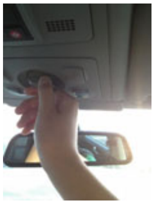
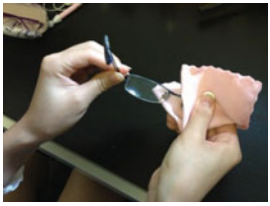
The two tasks in Table 13 second row appear almost identical, but the direction of motion is different in terms of hand coordinates. Twisting the key happens around y-axis of the hand (the axis out of the palm), and twisting the knob happens around the x-axis of the hand (the direction aligning with the forearm).

Table 12 Large diameter cylinder grasp examples

Example		
Force type	Hold	Hold
Motion dir	Around y axis (hand)	–
Force dir	–	−z (global)
Flow	Free motion/bound force	Bound force
Annotation	Drink water	Hold a bottle
Example		
Force type	Hold	Grab and press
Motion dir	x(45)y (hand)	–
Force dir	–	z (global)
Flow	Free motion/half bound force	Bound force
Annotation	Throw paper	Grab cabbage
Example		
Force type	Squeeze	Twist
Motion dir	–	Around z axis (hand)
Force dir	Inwards (hand)	Inwards (hand)
Flow	Bound force	Free motion/bound force
Annotation	Squeeze an empty soda bottle	Squeeze towel to dry

Some motions are in the same direction but with different force types and flow as shown in Table 13 third row. In this case, the force based interactions are both in the xy-plane of the hand (or equivalently the object), but one example has free motion while gently holding the grasped object and the other has motion relative to the object that is constrained to maintain forceful contact for cleaning. These differences are reflected in the differing annotations.

Table 13 Lateral grasp examples

Example		
Force type	Pull	Pull
Motion dir	−x (hand)	xz plane (hand)
Force dir	–	–
Flow	Bound motion/bound force	Half bound motion/bound force
Annotation	Put on gloves(along the arm)	Drag toilet paper
Example		
Force type	Twist	Twist
Motion dir	Around y axis (hand)	Around x axis (hand)
Force dir	–	–
Flow	Bound motion	Bound motion
Annotation	Twist the key to start up the car	Twist the knob in car
Example		
Force type	Hold	Rub/stroke
Motion dir	xy plane (hand)	xy plane (hand)
Force dir	–	Inwards (hand)
Flow	Free motion/half bound force	Half bound motion/bound force
Annotation	Give card to someone	Wipe glasses
Example		
Force type	Hold	Hold
Motion dir	z (global)/−z (global)/around x axis (hand)	Around x axis (hand)
Force dir	–	–
Flow	Free motion/bound force	Half bound motion/bound force
Annotation	Eat with spoon	Pour washing powder

5.3 New Types

From our observations, the existing taxonomy that served as our starting point [5] has covered many types of grasps. However, there exist some actions which are not represented by their taxonomy, for which we have created new categories in the database. Some of the new entries involve deformable objects. Some are very specific gestures such as opening a soda can and tying shoes. Overall, we have added 40 new categories. We illustrate 8 of them in Table 14. All classifications and annotations can be found in our database [14]. Some, but not all of the new grasp types can be found in other taxonomies, such as those of Kapandji [18] and Buckland et al. [4].

6 Discussion

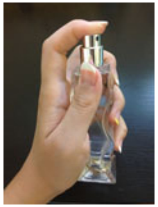
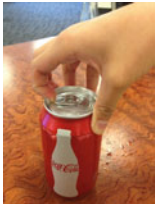
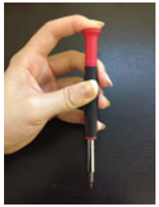
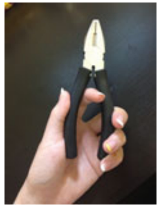
Effective grasp taxonomies capture not only hand shape, but also the nature of contact between the hand and object. The best in this regard is perhaps the Kamakura taxonomy [2], which illustrates in great detail regions on the hand that come in contact with the object. The patterns and extent of these regions reveals much, especially when considering grasp control and robot hand design.

However, we find annotating only shape and contact to be insufficient to convey important differences between everyday actions; in part because this set of actions is more broad than grasping, but also because many grasps that may look similar from a snapshot involve very different intentions—different uses of the hand to accomplish a task. We find that to communicate these differences, we need to express the type of force, directional information, and stiffness information for the action.

It is interesting to note the similarities between our annotations and the parameters required for impedance control [12] or operational space control [13], where one expresses a task in terms of the desired impedance or motion/force/stiffness properties of the manipulator. Annotations such as those we propose here could form the starting point for a learning-from-demonstration or coaching system where the user indicates to the robot coordinate frames and directions best suited for position control and force control, along with indications of the level of force or stiffness required for the task. In particular, we found the use of English language verbs very promising for conveying the type of force desired in a way that was intuitive for our subjects, and the use of multiple coordinate frames (hand, object, and world) make it easier to specify axes along which motion and force should be emphasized or constrained. It is of great interest to us to explore mechanisms for translating such annotations into robot controllers and allowing users to provide feedback to adjust those controllers in a language that is natural to them.

The similarities between our classification scheme and Laban Movement Analysis (LMA) [11] are also intriguing and invite further exploration. Perhaps we

Table 14 New type examples

Example				
Annotation	Tie	Shuffle cards	Lift up the switch	Scratch
Example				
Annotation	Press perfume bottle	Open soda bottle	Use screwdriver	Use pliers

may consider the static grasps of the conventional taxonomies as Shape Forms—static shapes that the hand may take while grasping an object. Annotation mechanisms within the category of Space may capture our intent when annotating motion and force directions, where we consider natural coordinate frames and landmarks that serve to orient the action. Annotation mechanisms within the category of Effort were motivating to us when considering how to discriminate between grasps. Although we did not make direct use of the Action Effort verbs (Float, Punch, Glide, Slash, Dab, Wring, Flick, and Press), many of them are represented in our force list of Table 1. In addition, we attempted to directly adopt the Effort category of Flow to allow users to discriminate between stiff and tightly controlled vs. free or flowing intent. We are interested to explore further how theory and practical experience from LMA may allow us to create more precise and comprehensive annotations.

Although there are similarities between our annotation scheme and LMA categories, there are also differences. For example, although our verb list is similar to the Action Effort verbs, there are verbs in our list that may fit one or more Action Effort verbs depending on how the action is performed. For example, in our database subjects used "Press" for forcefully supporting a cabbage for cutting and also for lightly pressing a small button, which may correspond to different Action Effort verbs such as "Press" and "Dab." In addition, there are items in our verb list that do not correspond well to the Action Effort verbs, such as "Put In" and "Take Out." The largest conceptual difference seems to be that our subjects considered verbs in our list to express *what* the hand was doing, as opposed to *how* the action was performed. Given this conceptual difference, it is interesting to see the level of similarity we do see in the two sets of verbs.

We also found that we needed to give our lightly trained users a great variety of verbs as options to specify force intent. We have listed 20 such verbs in Table 1 and have no doubt that a more extensive survey of everyday actions will require adding others. Intent of an action as it affects function and appearance of grasping appears to be challenging to capture and communicate in a manner that can discriminate between actions that are evidently different to both the performer and the observer.

One limitation of this database is that we need a more accurate system for describing the direction of motion and force that accommodates directions that do not perfectly align with an easily identifiable single axis. However, interestingly, this situation appears to be uncommon.

We can also ask whether all entries in our database are relevant for humanoid robots. We believe that as robots become more pervasive, especially in home, health care, and rehabilitation scenarios, a large majority of the grasps depicted here will become of interest. However, we did not attempt to make this distinction.

It may be possible to organize this database from a different point of view, such as making the force types or motion types the central classification rather than grasp type. We chose grasp type as the first level of organization in order to be consistent with existing taxonomies. However, it is interesting to consider whether a different organization may lead to a simpler or more intuitive way of describing these results.

Acknowledgement This research was supported by NSF awards IIS-1218182 and IIS-1344361. We would like to thank Daniel Troniak and Michael Koval for many helpful comments, discussion and suggestions.

References

1. C.L. MacKenzie, T. Iberall, *The Grasping Hand*, vol. 104 (Elsevier, Amsterdam, 1994)
2. N. Kamakura, M. Matsuo, H. Ishii, F. Mitsuboshi, Y. Miura, Patterns of static prehension in normal hands. Am. J. Occup. Ther. **34**(7), 437–445 (1980)
3. M.R Cutkosky, On grasp choice, grasp models, and the design of hands for manufacturing tasks. IEEE Int. Conf. Robot. Autom. (ICRA) **5**(3):269–279 (1989)
4. D.J. Buckland, S.J. Edwards, J.D. McCoy-Powlen, *Developmental and Functional Hand Grasps*, 1st edn. (Slack Incorporated, 2002)
5. T. Feix, R. Pawlik, H. Schmiedmayer, J. Romero, D. Kragic, A comprehensive grasp taxonomy, in *Robotics, Science and Systems: Workshop on Understanding the Human Hand for Advancing Robotic Manipulation*, June 2009
6. S. Ekvall, D. Kragic, Grasp recognition for programming by demonstration, in *IEEE International Conference on Robotics and Automation (ICRA)* (IEEE, 2005), pp. 748–753
7. G. Heumer, H.B. Amor, M. Weber, B. Jung, Grasp recognition with uncalibrated data gloves —a comparison of classification methods, in *IEEE Virtual Reality Conference, VR'07* (IEEE, 2007), pp. 19–26
8. A.M. Dolar, R.D. Howe, The highly adaptive SDM hand: design and performance evaluation. Int. J. Robot. Res. **29**(5), 585–597 (2010)
9. S. Bing Kang, K. Ikeuchi, Toward automatic robot instruction from perception-mapping human grasps to manipulator grasps. IEEE Int. Conf. Robot. Autom. (ICRA) **13**(1):81–95 (1997)
10. R. Wimmer, Grasp sensing for human-computer interaction, in *Proceedings of the Fifth International Conference on Tangible, Embedded, and Embodied Interaction* (ACM, 2011), pp. 221–228
11. J. Newlove, J. Dalby, *Laban for All* (Routledge, 2003)
12. N. Hogan, Impedance control: an approach to manipulation: Part III: implementation. J. Dyn. Syst. Meas. Contr. **107**(1), 8–16 (1985)
13. O. Khatib, A unified approach for motion and force control of robot manipulators: the operational space formulation. IEEE Int. Conf. Robot. Autom. (ICRA) **3**(1):43–53 (1987)
14. J. Liu, F. Feng, Y. Nakamura, N. Pollard, Grasp taxonomy in action online database: http://graphics.cs.cmu.edu/nsp/projects/graspDatabase/database.html (2014)
15. J. Liu, F. Feng, Y. Nakamura, N. Pollard, A taxonomy of everyday grasps in action, in *IEEE International Conference on Humanoid Robots (Humanoids 2014)*, November 2014
16. I.G. Schlesinger, Der mechanische aufbau der künstlichen glieder, in *Ersatzglieder und Arbeitshilfen* (Springer, Berlin, 1919), pp. 321–661
17. J.R. Napier, R. Tuttle, *Hands* (Princeton University Press, New Jersey, 1993)
18. I.A. Kapandji, L.H. Honoré. *The Physiology of the Joints: Annotated Diagrams of the Mechanics of the Human Joints*, vol. 1 (E.&S. Livingstone, London, 1970)
19. J.Z. Zheng, S. De La Rosa, A.M. Dollar, An investigation of grasp type and frequency in daily household and machine shop tasks, in *IEEE International Conference on Robotics and Automation (ICRA)* (IEEE, 2011), pp. 4169–4175
20. L.Y. Chang, N.S. Pollard, Video survey of pre-grasp interactions in natural hand activities, in *Robotics: Science and Systems (RSS) 2009 Workshop on Understanding the Human Hand for Advancing Robotic Manipulation*, June 2009

21. F. Worgotter, E.E. Aksoy, N. Kruger, J. Piater, A. Ude, M. Tamosiunaite, A simple ontology of manipulation actions based on hand-object relations. IEEE Trans. Auton. Mental Dev. **5** (2):117–134 (2013)
22. J.M. Elliott, K.J. Connolly, A classification of manipulative hand movements. Dev. Med. Child Neurol. **26**(3), 283–296 (1984)
23. I.M. Bullock, R.R. Ma, A.M. Dollar, A hand-centric classification of human and robot dexterous manipulation. IEEE Trans. Haptics **6**(2):129–144 (2013)
24. T. Torigoe, Comparison of object manipulation among 74 species of non-human primates. Primates **26**(2), 182–194 (1985)
25. B. Kopp, A. Kunkel, H. Flor, T. Platz, U. Rose, K.-H. Mauritz, K. Gresser, K.L. McCulloch, E. Taub, The arm motor ability test: reliability, validity, and sensitivity to change of an instrument for assessing disabilities in activities of daily living. Arch. Phys. Med. Rehabil. **78** (6):615–620 (1997)
26. C. Collin, D.T. Wade, S. Davies, V. Horne, The Barthel ADL index: a reliability study. Disabil. Rehabil. **10**(2), 61–63 (1988)
27. J.M. Linacre, A.W. Heinemann, B.D. Wright, C.V. Granger, B.B. Hamilton, The structure and stability of the functional independence measure. Arch. Phys. Med. Rehabil. **75**(2):127–132 (1994)
28. N. Pollock, M.A. McColl, A. Carswell, T. Sumsion, The Canadian occupational performance measure. Client-Centred Practice Occup. Ther. A Guide Implementation 103–114 (2006)
29. Y.S. Choi, T. Deyle, T. Chen, J.D. Glass, C.C. Kemp, A list of household objects for robotic retrieval prioritized by people with ALS, in *IEEE International Conference on Rehabilitation Robotics, ICORR 2009* (IEEE, 2009), pp. 510–517
30. J.R. Napier, The prehensile movements of the human hand. J. Bone Joint Surg. **38**(4):902–913 (1956)
31. A.A. Samadani, S. Burton, R. Gorbet, D. Kulic, Laban effort and shape analysis of affective hand and arm movements, in *2013 Humaine Association Conference on Affective Computing and Intelligent Interaction (ACII)* (IEEE, 2013), pp. 343–348

Laban Movement Analysis—Scaffolding Human Movement to Multiply Possibilities and Choices

Angela Loureiro de Souza

Abstract Rudolf Laban (1879–1958) and his colleagues, based on their transversal experience in different creative fields, proposed a complex approach of human movement that values its process of transformation, the plasticity of human beings, and the possibility of expanding human referential and movement repertoire. What is now called Laban Movement Analysis is composed of four main and interdependent points of view concerning movement: "Effort" deals with qualitative changes in the use of the motion factors Time, Space, Weight and Flow, their polarities, configurations and phrasings; "Shape" concerns the capacity of changing body shapes and its link with human adaptation to internal and external issues; "Body" focus on gesture, posture and their merging, and patterns of total body organization; "Space" speaks about personal, interpersonal and general space and reveals the spatial scaffold of movement (dimensions, planes and diagonals). The paper approaches some essential aspects of each domain, focusing on their connection and interdependence, and exposes the different notations proposed by the Laban Movement Analysis.

1 Introduction

Almost a century ago, different people in different places, with different backgrounds, questioned themselves about dance and movement, as well as their vital role in society. Among them, Rudolf Laban can be distinguished. He federated, around his ever-changing proposals, artists and students that were also in quest of another way of moving and creating. Irmgard Bartenieff, Warren Lamb, Lisa Ullmann, Albrecht Knust and Kurt Jooss make up part of the first generation that worked directly with him to settle the basis of this approach of movement.

A.L. de Souza (✉)
Movimento 3D Association, Paris, France
e-mail: angelaloureiro.cma@gmail.com

J.-P. Laumond and N. Abe (eds.), *Dance Notations and Robot Motion*,
Springer Tracts in Advanced Robotics 111,
DOI 10.1007/978-3-319-25739-6_13

Since the beginning of the 20th century till now, the work proposed by what may be called the "Laban Collective" has been enriched and, because it is grounded on creation, experimentation and observation, received the input of many artists—experimenters. In spite of this mobile frontier, the core of Laban's proposals remained and guaranteed the coherence of this approach of movement. It is question of considering movement as a process of changing body and shape configurations, spatial positioning, different quality usage. It is question of discovering patterns and structures that will order this process of transformation, allowing its observation. And it is also question of plasticity and adaptability, of multiplying movement possibilities and of creating nuances and shades.

Nowadays, four main interdependent fields of exploration and observation can be distinguished: Body, Effort (named before Eukinetics), Shape and Space (named also Choreutics). Each one must be considered as a point of view to approach the complexity of movement. Together they form what is called Laban Movement Analysis.

2 Laban Movement Analysis: Body, Effort, Shape, Space

2.1 Body

The point of view of the "Body" is based on a global and mobile way of considering the body. Rudolf Laban conceives body architecture in relation to spatial pulls and zones, its three-dimensional aspect being stressed. Exertion and recuperation, stabilizing and mobilizing forces, frequent themes in Laban's approach, denote a non-static conception of the body. This general approach is enriched by the importance given to the flow of movement, influenced by the way different parts of the body are set in motion. In his book The Mastery of Movement (1950), R. Laban studies body actions (directions, shapes, rhythmical development, organization of phrases), body parts and gestures, as well as their relationship with the environment. These ideas are the basis of the proposals of Irmgard Bartenieff (1900–1981) through her "Fundamentals", and of Warren Lamb (1923–2014) through his "Movement Pattern Analysis". They clarify the role of the alternation and merging of postures and gestures, the importance of initiation and sequencing of movement, of internal body connections and spatial intent, of patterns of total body organization through a developmental perspective:

- Gesture is defined as a movement of a body part, posture as a movement that engages the totality of the body. The study of gesture will establish if body parts are acting simultaneously, sequentially or successively. The study of posture will explore actions in which the whole body is engaged, as changing support, spiraling, running, turning, falling, etc. Warren Lamb singles out the point or the phrase in which the posture merges into a gesture, or the gesture into a posture. What he calls PGM (Posture Gesture Merger) is part of the individual way of moving, of its contrived or integrated movements.

- Initiation and sequencing of movement deals with the part of the body that begins an action (does it begins distally, centrally, with the lower body, with the upper body?) and with the way movement spreads through the body.
- Body connections recreate the possibility of sequencing movement. They have been addressed in the "Bartenieff Fundamentals", movement sequences that engage the person deeply in the path of mobility. They not only deal with the link between body parts through muscles and ligaments, but with their "connectedness", defined as the capacity of allowing the movement impulse to pass through the body. Body connections are also supported by "spatial intent": clarifying direction organizes movement both in its functional and expressive aspects.
- Patterns of total body organization, equally addressed by the "Bartenieff Fundamentals", study fundamental structures through a developmental perspective: connections between the center and the periphery of the body in a radiation pattern; between the head and the coccyx; between the lower and the upper body; between left and right sides of the body; between the four quadrants of the body, in a contralateral pattern. Movement sequences based on these patterns, as well as their approach through a creative and expressive practice, revisit the path each human follows to acquire verticality, from lying to uprightness. Incorporated through experience in the first year of life, patterns enhance differentiation and connection between body zones and spatial zones, and through them between the person and his environment. They are one of the basis of human mobility and expressivity.

2.2 Effort

The theme of movement dynamics appeared in Rudolf Laban's researches and artistic practice since the twenties. First called Eukinetics, the study of the qualities of movement later received the name of Effort. During the forties, four books addressed this particular point of view: Effort (1947) addresses these qualities in the workplace and in industry; Modern Educational Dance (1948) proposes a movement education in school based on the understanding of movement's specificity of the different childhood groups of age; The Mastery of Movement (1950) studies body parts in relation to energy usage in dance, theatre and mime; Effort and Recovery—seventeen studies of people in motion (1950—unpublished) addresses the important question of movement and its recovery.

The "Effort" point of view focuses on the link between intention and motivation with different qualities of movement. Rudolf Laban composed a chart of qualities, which he named "motion factors": Weight, Flow, Time and Space. Each one is composed of two polarities in a range, called elements.

The Flow factor deals with the way each one controls the progression of movement and is composed of a Free and a Bound Element. It is considered as the basic factor, from which the others will emerge.

The Weight factor deals with the way each one, in different situations, relates to gravity in order to accomplish a specific intent. It is composed of a Light and a Strong Element.

The Time factor deals with the way each one relates to time, wishing or being obliged to accelerate or decelerate movement, no matter the duration. This factor is composed of a Sudden and a Sustained Element.

The Space factor deals with the way you give attention, whether you focus one point or several points. It is composed of a Direct and an Indirect Element.

These factors don't appear alone; they combine to create multiple configurations composed of two (named States), three (named Drives) or four polar elements (named Full Effort). They are frequently transitioning from one to another, creating rhythms and phrasings. For example, a movement characterized by the combination Light Weight and Indirect Space will be completely different of another composed of the combination Light Weight, Free Flow and Direct Space. Both configurations can appear in a sequence of movement, creating a specific phrasing.

Effort deals with the way a movement is performed and varies in function of specific intents and contexts. Are we the same when we turn vigorously or lightly, when we hurry to take a bus or when we stroll without paying attention to the world, when we assemble pieces of an ancient clock or when we do finger-painting, when we point out someone or when you address yourself to everyone in a room? The chart of qualities is like a color chart: we can multiply nuances and ways of being.

Rudolf Laban alludes to the existence of combinations of Effort elements in a particular rhythm, frequently very personal, that he names movement phrases. For example, a sequence can begin with movements colored by Strong Weight and Indirect Space, evolves keeping the Strong Weight and replacing Indirect Space for Direct Space, and concludes substituting Strong Weight for Light Weight. The resulting rhythm of the sequence has its origin in this dynamic modulation (Fig. 1).

Further work developed the concept of Phrasing: the sequencing of movement qualities produces specific intensities, patterning of different energies that will change the way movements are done. Warren Lamb, Marion North (1925–2012) and Vera Maletic (1928) studied in depth the phrasing aspect of movement, the first in relation to work relationships, the second in therapeutic context and the third in choreography and pedagogy. Eight types of phrasing were named: Even (when the same intensity is maintained), Increasing-Intensity (when intensity gradually increases), Decreasing-Intensity (when initial intensity diminishes gradually), Increasing–Decreasing Intensity (when energy grows and diminishes without a culminating point), Decreasing–Increasing Intensity (energy diminishes then increases), Accented (repeated accents), Vibratory (a series of quick actions) and Resilient (rebounding movements). In any of these contexts, Phrasing is seen as

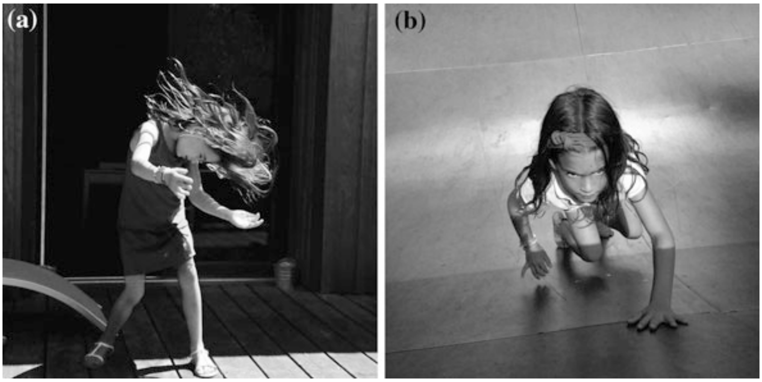


Fig. 1 **a** Her dance combines free flow and sudden time. **b** Her determination is supported by direct space and strong weight

fundamental to determine moving styles. The access to different ways of combining Effort elements opens many possibilities of responding to internal and relational issues.

2.3 Space

In his book Choreutics, published posthumously in 1966, Rudolf Laban develops the space-movement aspect: the moving human being would be like a living architecture, recreating his balance all the time, activating mobilizing and stabilizing forces, changing internal cohesion with a tensegrity perspective, creating through shapes in space a three-dimensional whole.

Looking at movement from this point of view, different kinds of space will be named: general, interpersonal and personal, each one representing an orientation system of reference to individuals.

- General space is the place where the person is located: it is invested through straight and curved pathways, is structured through directions, and has a center, a mid-space and a periphery.
- Interpersonal space is the all-around distance between people, and varies substantially in function of personal and cultural issues.
- Personal space has a specific name: the Kinesphere, a term coined by Laban to define the volume created and occupied by each person (Fig. 2).

As the general space, the Kinesphere has a center, a mid-space and a periphery, that is differently invested by each person. The center corresponds to the center of the body, the periphery is what defines its limits, and the mid-space is the area between the surface of the body and its periphery. Depending on the movement and on the mood of the mover, the kinesphere can change its size, becoming bigger or smaller.

Fig. 2 He is at the center of his personal space, the Kinesphere

Its limits are not impermeable, but are conceived as an interface of exchange between the person and the environment. The Kinesphere has spatial zones: up and down, back and front, right side and left side; it has levels: high, middle, low; and is structured by dimensions, planes and diagonals. Dimensions compose the basic scaffold of the Kinesphere, each one presenting two opposite directions: the vertical dimension is composed by high and low; the sagittal dimension by front and back; the horizontal dimension by right side and left side. Their intersection point corresponds to the central zone of the body. The mover explores the main axes of the body. The three planes—horizontal, vertical and sagittal—will open the possibility of more complex movements: each plane is the result of the combination of a primary and a secondary spatial tension. When they are put together, their intersection point corresponds to the center of the body. The person experiments movements of the spine that will challenge his balance: inclination (vertical plane), flexion/extension (sagittal plane), and rotation (horizontal plane). Another spatial structure will challenge even more the plasticity and balance of the human body: in the zones created between the planes, diagonals combine three spatial tensions in equal proportion. The four diagonals cross the body through its center going up and down, to the right and left side, forwards and backwards, combining these spatial pulls: for example, one diagonal will have a segment that goes up, forward and to the right side, and an opposite segment that goes down, backwards and to the left. The mover experiments torsions and spirals, challenging transformations of his body's shape and movement. Dimensions, planes and diagonals intersect in the center of the body and structure human movement in the Kinesphere.

Different people do not invest the spatial scaffold of the Kinesphere the same way. Practitioners/researchers, as Irmgard Bartenieff, Warren Lamb, Carol-Lynne Moore and Ellen Goldman, have shown in their studies how much the geometry of movement is embedded in child development, in communication and even in the process of making decisions. Different styles of moving and of being can be perceived through the spatial point of view, that cannot be separated from the body. One of the bases of the Laban Movement Analysis is the "spatialized" body, structured by directional pulls, and always dealing with balance and unbalance (Fig. 3).

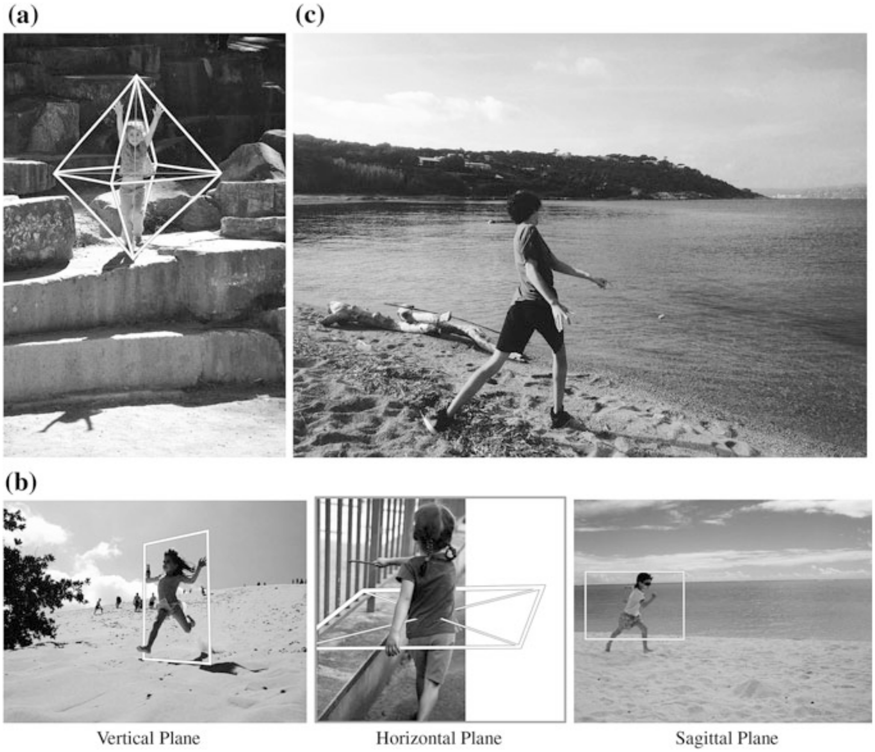


Fig. 3 **a** The 3 dimensions inside the octahedron. **b** The 3 planes. **c** Diagonal spatial tension

2.4 Shape

The plasticity of the human body, conceived as the capacity of changing volume and shape, is an integrative and vital point of view: the body changes shape in an inner oriented or in an outer oriented manner in order to cope with different levels of need and with different kinds of relationship with the environment. Since birth human beings must develop the capacity of changes in volume, first due to breath, to seeking comfort, to avoid discomfort, to attraction and to repulsion. The fundamental and tridimensional movement of growing and shrinking defines this period. Later, it is observable that the body follows six main processes of changing shape: rising and sinking/lengthening and shortening; advancing and retreating/bulging and hollowing; spreading and enclosing/widening and narrowing.

Multiple combinations result from these fundamental elements: for example, a movement can easily combine a retreating with a sinking shape, or a retreating with an ascending one; another combination can be the result of a spreading combined with an advancing process. Each resulting shape of the moving body will have

Fig. 4 Body shapes change in relation to intention

specific origins and impacts in the mood of the mover and in his relations with the environment. These multiple possibilities create multiple responses to inner and outer stimuli.

Shape interweaves with Body, with Space and with Effort. A monolithic body cannot change shape; differentiation and connection of body parts are at the same time the condition and result of body plasticity. Spatial pulls are present in each of the shape changing processes: rising and sinking/lengthening and shortening processes are related to verticality; advancing and retreating/bulging and hollowing processes are related to sagittality; spreading and enclosing/widening and narrowing processes are related to horizontality. Effort factors and elements interweave with body shapes in the sense that a certain shape will facilitate an action colored by a specific Effort element. In the document "Effort-Shape analysis of movement—the unity of function and expression", Irmgard Bartenieff announces the hypothesis that a complex of biological factors (as the alpha-gamma system, postural reflexes, the senses, the basal ganglia) could explain these affinities; nevertheless, as activities and expressions become more complex they would appear less consistently (Fig. 4).

3 Movement Scales as Models of Coherence

Rudolf Laban associated dimensions, planes and diagonals to specific polyhedrons: dimensions to the octahedron, planes to the icosahedron, and diagonals to the cube. They fit into each other and are the references to "movement scales", cyclic sequences conceived as models to explore zones, levels and directions of the Kinesphere, to amplify our reach space, and to integrate space variations in the search of balance. These scales offer the mover a strong experience of the mobilizing and stabilizing forces that are always present in human movement.

The scales also develop the coherence and integration of spatial structures, Effort factors and Shapes. Rudolf Laban observed in daily life, in artistic, working and fighting contexts, that a specific direction was easily supported by a certain quality of movement and by a certain shape. The first attempt to create a model articulating all these physical realities in a movement sequence was the Dimensional scale. It became the referential model to the other scales.

The mover will first explore the vertical dimension, then the horizontal and after the sagittal:

- Going high (vertical dimension) with Light Weight, this movement is supported by the rising/lengthening process.
- Going down (vertical dimension) with Strong Weight, this movement is supported by the sinking/shortening process.
- Going sideways (horizontal dimension) with a side of the body that crosses its midline with Direct Space, this movement is supported by the enclosing/narrowing process.
- Going sideways (horizontal dimension) with a side of the body that opens, this movement is supported by the opening/widening process.
- Going back (sagittal dimension) with Sudden Time, this movement is supported by the retreating/hollowing process.
- Going forwards (sagittal dimension) with Sustained Time, this movement is supported by the advancing/bulging process.

The other scales follow this chart of affinities; their spatial structure being more complex, the resulting movement will also be more complex and will constantly challenge our balance. Taking for example the Diagonal scale and choosing just one of the diagonals that compose the sequence, it is possible to feel and see this complexity. A triad of directions composes the diagonal, with opposite spatial pulls. If we choose the diagonal that goes from front/high/right to back/down/left, the resulting movement will ally advancing-bulging/rising-lengthening/opening-widening processes (Shape) with Sustained Time-Light Weight-Indirect Space (Effort). Then the mover must completely change the cluster of affinities to address himself to the opposite pull and to ally: retreating-hollowing/sinking-shortening/enclosing-narrowing processes (Shape) with Sudden Time-Strong Weight-Direct Space (Effort). The mover will work not only the contrasting movements but the transition between them: from a floating one (the movement resulting from the first combination) to a punching one (the movement resulting from the second one).

The scales are models that engage the whole body in changing Body, Effort, Shape and Space configurations and, through this workout, they multiply one's possibilities of expression and function.

4 Daily Life as Choices of Coherence

Script 1: You are in the middle of a crowd and you are looking for someone; you see him but he does not see you. In order to call his attention, you elongate yourself and make big gestures. Then you try to make your way among people and you succeed in getting closer. You open your arms to embrace your friend and together you turn till you stop.

This script can be realized in different ways, depending on the people that are engaged in the action. The following movement analysis is one of these possibilities: in the beginning you phrase the Space factor, the indirect element moves toward the direct element. It will be easier to activate this factor if you rise and spread your body. This shape will support the free flow of your big gestures. Then you will probably have to enclose and narrow your body to move among people, certainly adding the Time factor. The arms opening gesture, that can be done with free Flow and Sudden Time, is supported by a spreading and widening shape process, followed by an enclosing process that supports the movement of embracing, when the Strong element of the Weight factor will appear. Then you and your friend activate the Flow factor in order to turn (Free Flow) and then stop (Bound Flow).

Script 2: A young child is playing on the floor. You address her and eventually play with her.

This script can be realized in different ways, depending on your way of addressing and making contact with the child. One possibility: you decide to get closer quietly, you lower yourself smoothly to put yourself on the same space level, and you speak softly and make delicate gestures. The movement analysis would be: in the beginning, your movements are in Bound Flow combined with Sustained Time and Direct Space; then the shape of your body changes, you can keep the same Effort configuration; the total shape remains and supports gestures done with Light Weight and modulations of Flow.

These two examples shows how humans are constantly changing configurations, in order to respond to inner and outer stimuli, to adapt to others and to the environment by modulating their movement choices.

5 Notations

Kinetography Laban, named in the Anglo-Saxon tradition Labanotation (subject developed in the chapter written by Jacqueline Challet-Haas), notates movement in its development and transformation. The scores are precise and allow the reader and the mover to follow the continuity of movement in its bodily, spatial and temporal structure.

Together with Kinetography, Laban Movement Analysis proposes other notational supports. They approach movement through different angles and don’t focus on the precision of each movement, but in the mood of a sequence, in its phrasing or in the way body shape changes. The choice of the notation depends on the proposal of the research worker, the choreographer, the dance teacher, the dance therapist, etc. For example, in a therapeutic context it can be more significant to register that someone is predominantly moving in a light and indirect way (Light Weight and Indirect Space, a combination of Effort elements), with an Even

Phrasing (no variations), with a restricted quality repertoire, than to notate specific gestures. In an artistic context, it may be enlightening to point out the predominance of a specific Effort configuration, as Passion Drive (when Weight, Flow and Time factors combine to create movements that don't pay attention to what is out there—in this case, Space factor is latent). This configuration and its nuances, if it colors almost all movements done in the choreography, will create a very specific bodily state and an artistic statement full of emotion (the famous final of Pina Bausch's Rite of Spring is a perfect example). In an ethnographic context, it can be interesting to examine the consistency between community dances, work and fighting gestures: have they the same Effort elements, the same Phrasing? Or are they totally inconsistent?

In order to cope with these different needs, Laban Movement Analysis proposes an Effort notation, a Shape notation and a Phrasing notation. Composed of horizontal and vertical lines that cross in their middle, what is called Effort Grid organizes the factors and their elements in a simple notational structure. The small oblique line, named Effort accent, has two important functions: the vertical and horizontal lines need the oblique accent to become an Effort sign; if prolonged, it distinguishes the qualitative world of indulging (Light, Indirect, Free, Sustained) movements from the fighting (Strong, Direct, Bound, Sudden) ones. The score that results from this grid has no precision about a specific gesture or posture. It indicates the tone, the color, the mood (Fig. 5).

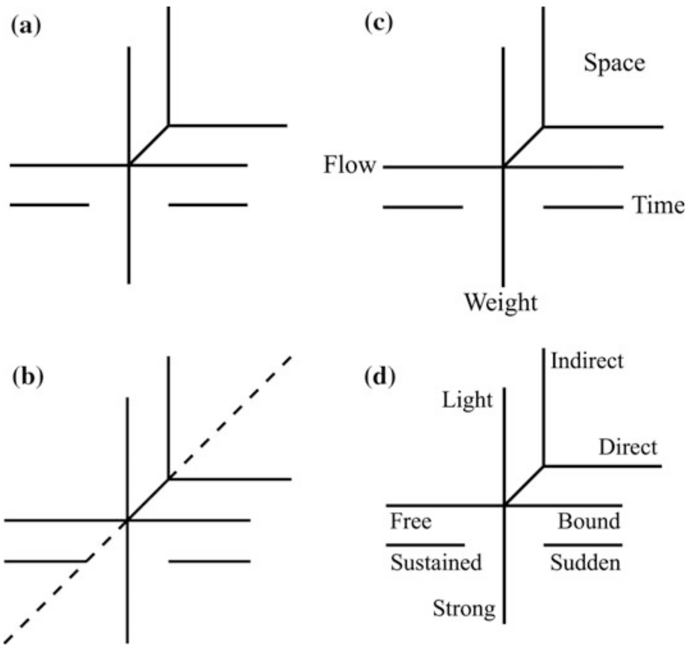


Fig. 5 **a** The effort grid. **b** *Upper left half* indulging elements. *Lower right half* fighting elements. **c** Effort factors. **d** Effort elements

Fig. 6 Punching

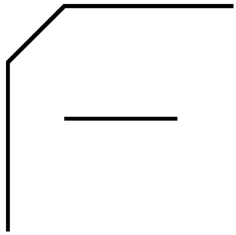
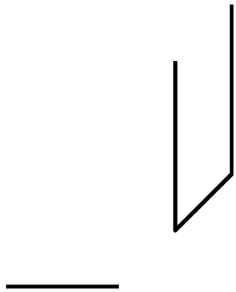


Fig. 7 Floating



Depending on the movement qualities of an event, different grids will be designed. For example, if the notated sequence is composed of Strong Weight, Direct Space and Sudden Time movements, similar to a punch, the resultant grid will be as in Fig. 6.

If another one is composed of Light Weight, Indirect Space factor and Sustained Time movements, similar to floating, the resultant grid will be as in Fig. 7.

Phrasing depicts Effort transformation in time, and also has a notational support, that combines aspects of Kinetography, of Effort notation and specific signs. The following illustration is an example of this notation: the vertical line indicates the duration of the sequence, the horizontal lines indicates when it begins and when it ends, the signs of increasing and decreasing are placed parallel to the main score. The indication of movement qualities and how it changes is placed inside these signs (Fig. 8).

Shape notation derives from the Effort Grid, and emphasizes the affinities between Effort and Shape. The main difference lies in the central zone: two inclined lines (in the Anglo-Saxon tradition), or a black dot in the intersection point (in the French proposal), replace the simple Effort accent (Fig. 9).

The different notations are not in opposition; they just assume that sometimes it is relevant to use one or the other, depending on what is being valued and why it is most significant to point the precision, the mood, or the plasticity. They can also be completely complementary; if necessary, a Kinetography score can be enriched by another type of score. Notations are conceived to facilitate movement analysis, but also to enhance movement practice and movement sharing. A score is the result of an analytical or creative purpose, and the beginning of multiple experiences and dialogues.

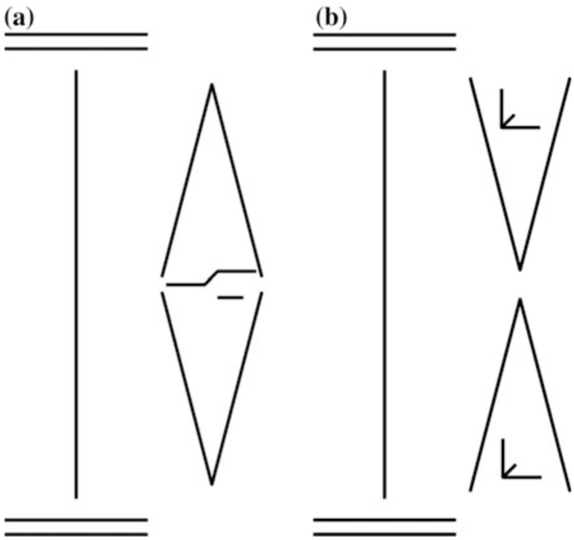


Fig. 8 **a** Example of increasing then decreasing intensity phrasing. A movement becomes more and more Direct (Space factor), Free (Flow factor) and Sudden (Time factor). Then, its intensity diminishes. **b** Example of decreasing then increasing intensity phrasing. A Light (Weight factor) and Bound (Flow factor) movement loses its intensity and then recovers it

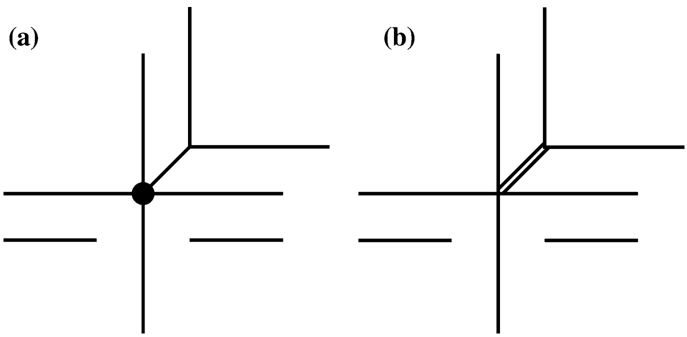


Fig. 9 **a** The shape graph (French version). **b** The shape graph (Anglo-Saxon version)

6 Conclusion

The fields of Body, Effort, Shape and Space are certainly specific angles of experimenting and of observing movement, but their relationship must be taken into account when analyzing movement in its complexity and in its process of change. Interested by any kind of human movement, Laban Movement Analysis proposes a study of its basic elements in order to understand not only what the movement is made of, but how its components are put together and how they sequence. LMA is

Fig. 10 Five different body configurations in response to the same event

at the same time concerned by the specificity of individual movement and by what can be called the alphabet of movement, basic elements and patterns that we all share.

The challenge would be the one of experimentation, of changing points of view, of enlarging movement possibilities, of multiplying modulations and intensities, responses and suggestions. Irmgard Bartenieff chose the expression "coping with the environment" for the title of her book. Isn't it the core of this approach of human movement? (Fig. 10).

Acknowledgments The photographs are courtesy of David Proux; Alice Proux realized the graphic design.

References

1. I. Bartenieff, D. Lewis, *Body Movement—Coping with the Environment* (Gordon and Breach Science Publishers, New York, 1980)
2. I. Bartenieff, M. Davis, *Effort-Shape Analysis of Movement—The Unity of Expression and Function* (Albert Einstein College of Medicine, Yeshiva University, New York, 1965)
3. P. Hackney, *Making Connections—Total Body Integration through Bartenieff Fundamentals* (Gordon and Breach Publishers, Amsterdam, 1998)
4. E. Goldman, *As Others See Us—Body Movement and the Art of Successful Communication* (Gordon and Breach Science Publishers, Lausanne, 1994)
5. E. Goldman, *The Geometry of Movement—A Study in the Structure of Communication*, 2nd edn. (Copyright Ellen Goldman, 2008)
6. R. Laban, *Effort*, 2nd edn. (MacDonald & Evans, Plymouth, 1974)
7. R. Laban, *Modern Educational Dance*, 2nd edn. (MacDonald & Evans, London, 1963). French edition: La Danse Moderne Educative, 2003, Ed. Complexe et Centre National de la Danse. Reediting: Ressouvenances, 2013, Villers-Cottêrets
8. R. Laban, *The mastery of movement*, 3rd edn. (MacDonald & Evans, London, 1971). French edition: La maîtrise du mouvement, 1994, Actes Sud, Arles
9. R. Laban, *Choreutics*, new edn. (Dance Books, Alton, 2011). French edition: Espace Dynamique, 2003, Nouvelles de Danse, Contredanse, Bruxelles

10. W. Lamb, *Posture and Gesture—An Introduction to the study of Physical Behavior* (Gerald Duckworth & Co., Ltd., London, 1965)
11. A. Loureiro de Souza, J. Challet-Haas, Les exercices fondamentaux de Bartenieff – une approche par la notation Laban. Ressouvenances, Coeuvres-et-Valséry (2008)
12. A. Loureiro de Souza, *Effort: l'alternance dynamique dans le movement* (Ressouvenances, Villers-Cottêrets, 2013)
13. V. Maletic, *Body, Space, Expression—The development of Rudolf Laban's Movement and Dance Concepts* (Mouton de Gruyter, Berlin/New York/Amsterdam, 1987)
14. D. McCaw, *An Eye for Movement—Warren Lamb's Career in Movement Analysis* (Brechin Books Limited, London, 2006)
15. C.L. Moore, K. Yamamoto, *Beyond Words—Movement Observation and Analysis* (Gordon and Breach Science Publishers, Pennsylvania, 1988)
16. M. North, *Personality Assessment through Movement* (Northcote House Publishers Ltd., Devon, 1972)
17. V. Preston-Dunlop, *Rudolf Laban—An Extraordinary Life* (Dance Books, London, 1998)
18. V. Preston-Dunlop, *A Handbook for Dance in Education*, 2nd edn. (MacDonald & Evans, London, 1980)
19. V. Preston-Dunlop, *Point of Departure: The Dancer's Space* (Lime Three Studios, London, 1984)

Benesh Movement Notation for Humanoid Robots?

Eliane Mirzabekiantz

Abstract Benesh Movement Notation (BMN) is a written system for analysing and recording human movement. It is a flexible tool that reduces three-dimensional body positions and actions in space over time to a series of two-dimensional key frames. Created in the twentieth century, BMN has been applied to fields as diverse as dance, gymnastics, mime, circus performance, anthropology, ergonomics, neurology, and clinical research. Might it also contribute to research in humanoid robotics? The intention of this paper is to provide the scientist with an introduction to its application across a variety of fields as well as a rudimentary understanding of the Benesh system, so that he may evaluate its potential contribution to robotics research. To that end, this paper explains how BMN conceptualizes movement and provides examples that illustrate how those fundamental concepts have been modified for special purpose projects. Given its demonstrated adaptability, the author is optimistic that the system may be extended through close collaboration between the notation expert and the robotics researcher.

1 The Genesis of Benesh Movement Notation (BMN)

In 1955, Dame Ninette de Valois, dancer, teacher, choreographer and director of the Royal Ballet, announced at a press conference the adoption of the Benesh Movement Notation to record its repertoire as well as the teaching of system at the Royal Ballet School. Soon after, in 1958, BMN was included among the technical scientific discoveries in the British government pavilion at the Brussels world expo.

It was in 1947, that Rudolf Benesh, an accountant and artist with a deep interest in scientific subjects, and his future wife, Joan Rothwell, a dancer with the Royal

E. Mirzabekiantz (✉)
Conservatoire National Supérieur de Musique et de Danse de Paris,
209 Avenue Jean Jaures, 75011 Paris, France
e-mail: e.mirzabekiantz@cnsmdp.fr

J.-P. Laumond and N. Abe (eds.), *Dance Notations and Robot Motion*,
Springer Tracts in Advanced Robotics 111,
DOI 10.1007/978-3-319-25739-6_14

Ballet, first considered the problems of devising a practical and efficient notation system. From then on, Joan and Rudolf started eight years of collaborative development. "Rudolf quickly set his mind to the problem, directly and indirectly drawing on concept of music, perspective drawings, linguistics and the new scientific disciplines of ergonomics, information theory and cybernetics." [1, p. 139].

With the appointment of the first Benesh choreologist[1] at the Royal Ballet in 1960, the use of BMN by choreographers expanded to the Commonwealth companies, Scandinavian, Germany, then for a short time to the US, and later to France in 1992. Thereby a wide range of choreographic scores has been compiled in different movement styles. The application of BMN to circus aerial acrobatics is just one example of its versatility [2, 3].

Outside the field of professional dance, BMN has been used in various research studies. In ergonomics, it has been used to analyse the movement of operators in front of machinery. In clinical and medical research, it has been used to analyse gait and record the movement of cerebral palsy patients. In anthropological studies, it has been applied to dance traditions of indigenous peoples in Australia and Africa. The system has also been considered for basic animation [4–7] and more recently for Human Computer Interaction [8].

2 How BMN Works [9]

In devising BMN, Rudolf Benesh aimed to create a comprehensive notation system capable of recording all forms of human movement. "Notation is a tool for creative thinking in research and composition. To be a creative tool the notation had to be unlimited in possibilities, and only a pure movement approach would satisfy this need." (Rudolf and Joan Benesh 1977, p. 6).

2.1 The Concept of Movement

Inspired by the chronophotography of Etienne-Jules Marey, one of the pioneers of animated photography, Rudolf retained the idea of capturing movement by a succession of "key frames" and tracing the path of movement using simple lines to summarise the intermediate positions (Fig. 1).

[1]Benesh choreologist: A person who is qualified in BMN and is employed to notate and revive works from a choreographic score.

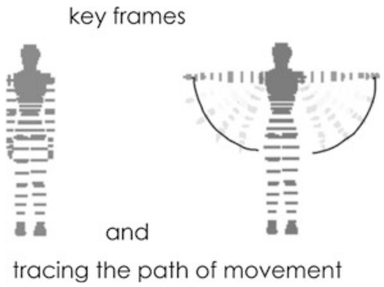


Fig. 1 Concept of movement retained by Rudolf Benesh

2.2 Transcription of the Concept

Looking for the most efficient representation of the body, Rudolf referred to Leonardo da Vinci's *Vitruvian Man* and drew five horizontal lines over the human figure (Fig. 2 to the left). He thus adopted a stave similar to the musical staff that reads from left to right, and organized the stave in order to easily identify three key elements: time (specified above the stave), body movement (specified within the stave), and spatial orientation (specified below the stave).

This organization of the stave facilitates the analysis and description of postures in time and space, and enables readers to integrate the three elements as a whole.

2.2.1 Body Movement

The five-line stave forms an ergonomic matrix on which the human body is projected. Figure 3 illustrates the anatomical landmark at which each line intersects the body. The stave becomes a scale of height.

BMN reduces the human figure to its essentials by using distinctive signs to locate extremities, joints and segments on the stave. Each frame then becomes a simple pictogram from which the reader extrapolates the whole body position (Fig. 4).

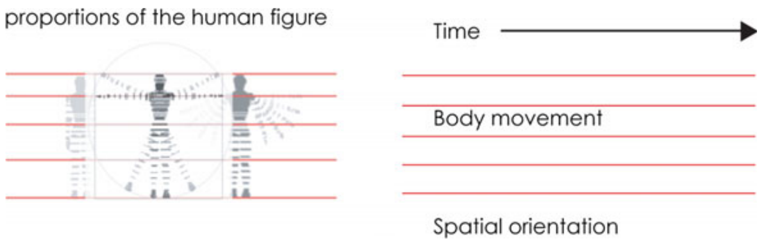


Fig. 2 To the *left*, the five line stave in reference to da Vinci. To the *right*, organisation of the stave where the 3 elements—time, body movement and spatial orientation—have their specific placement

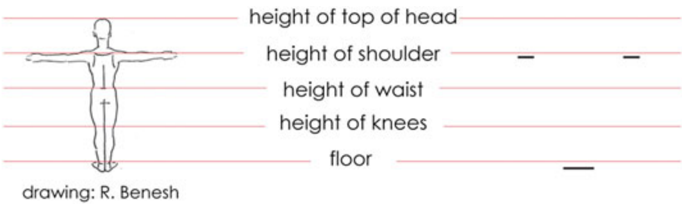


Fig. 3 The stave plots the hands just below shoulder height

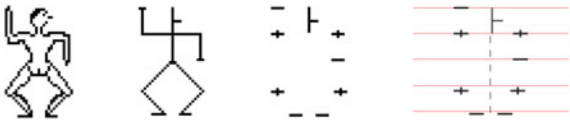


Fig. 4 The body representation is reduced to a few distinctive signs placed on the stave

Each frame is divided into left and right halves, so that the reader, viewing from behind, can easily identify left and right, i.e., the figure's left corresponds to the reader's left (see Fig. 5).

In order to plot a three-dimensional image on a two-dimensional page, the depth dimension is represented by three differently shaped signs that depict an extremity in front, level with, or behind the body (Figs. 6, 7 and 8).

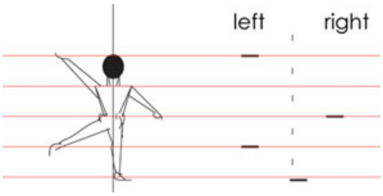


Fig. 5 Body viewed from behind. Standing on right foot, left foot to the side at knee height, both hands to the side, left hand at top of head height and right hand at waist height (The *dotted centre line* is a visual aid, not part of the notation)

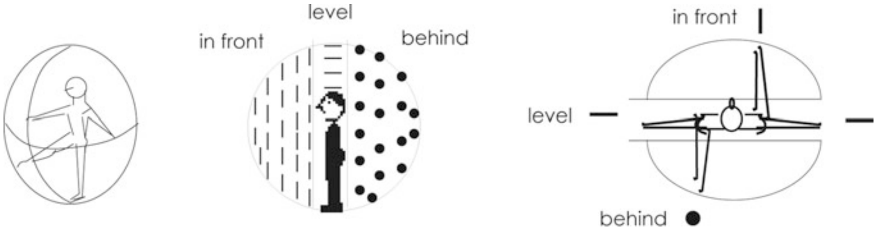


Fig. 6 Three peripheral spaces "in front of, level with, behind the body"

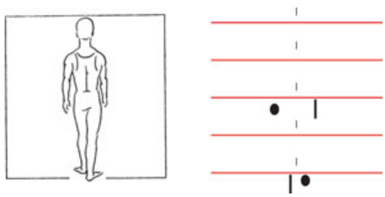


Fig. 7 Walking attitude. Standing on both feet (flat), left foot in front and right foot behind the body; both hands just below waist height, left hand behind and right hand in front of the body

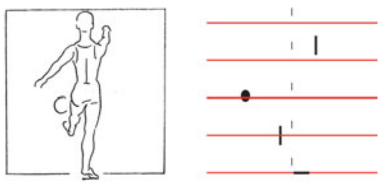


Fig. 8 Kicking a ball. Standing on the right foot, heel off the ground, left foot in front at knee height (as if having just kicked the ball) right hand in front, just above shoulder height, and left hand diagonally behind at waist height

These three signs that represent the extremities (feet and hands)—a vertical stroke, a dash and a dot—are the foundation of the evolving BMN alphabet. For example, these three basic signs are modified to a cross to identify the joints (knees and elbows) (Fig. 9).

Once the position of the limbs is recorded, the path of the extremities and the transitions from one key frame to another are shown. Using movement lines and locomotion lines, as shown in Fig. 10, all intermediate positions are summarized in an efficient visual manner.

Fig. 9 The cross signs are used to plot bent knees and elbows on the five-line stave

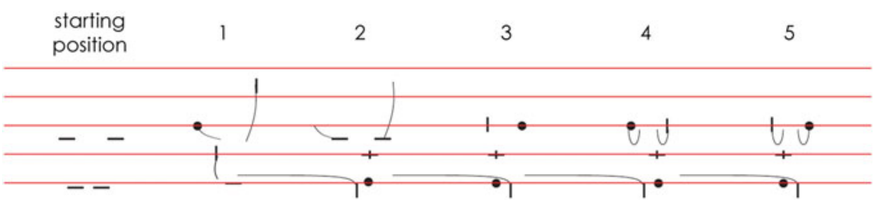


Fig. 10 A kick of an imaginary ball (on *1*) followed by four steps going forward

2.2.2 Movement in Relation to Time

BMN records a sequence of key frames. To specify rhythm and timing, signs are written above the stave as needed to show main beats and sub-beats. A key frame corresponds to a pulse beat unless otherwise specified. In Fig. 10, the sequence consists of a starting position followed by five movements executed on five regular beats.

The placement of a rhythm sign above a frame modifies the moment in time at which the position is reached. In Figs. 11 and 12, the "an" sign, corresponding to the half-beat, has been placed differently, which changes the rhythm of the sequence (Fig. 13).

2.2.3 Movement in Relation to Space

Direction faced, path of travel, etc., are shown by signs placed below the stave, as illustrated in Figs. 14 and 15.

To show direction faced in relation to the working area, a direction sign is placed below a key frame. The direction sign may be thought of as an arrow in which the tip has been reduced to a dot. The travel sign is a modified arrow that traces the path of travel.

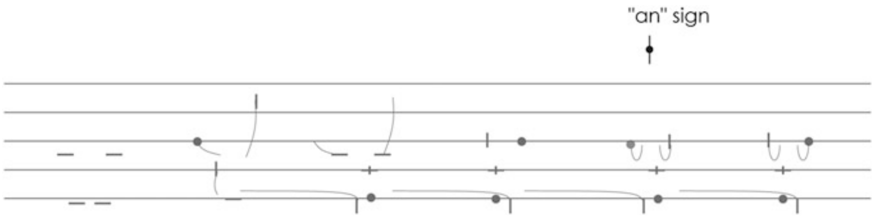


Fig. 11 Same movement sequence as in Fig. 10, with a different rhythm (1, 2, 3, "an" 4)

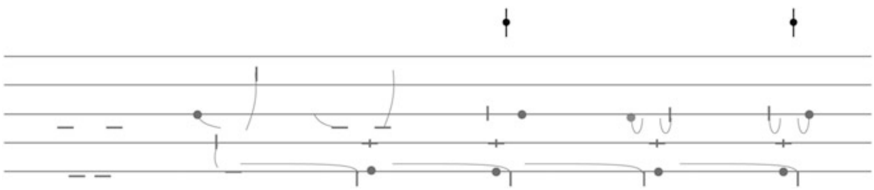


Fig. 12 As Fig. 10, with another rhythm (1, 2, "an", 3, "an")

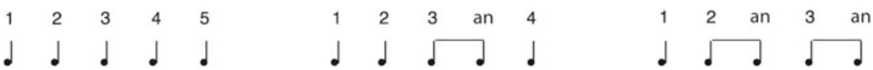


Fig. 13 Musical transcription of Figs. 10, 11 and 12

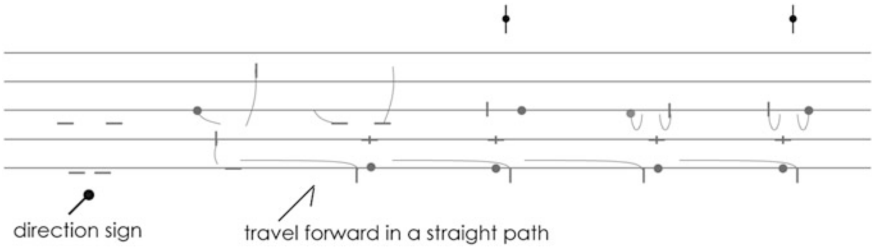


Fig. 14 Starting facing right front corner, the sequence will travel forward on a straight path

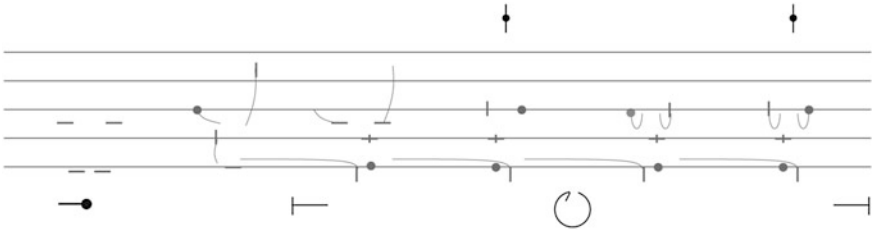


Fig. 15 Starting facing right wall, the sequence will travel forward in a small circle, clockwise

2.2.4 Details of Movement

The development of BMN for recording movements of eyes, gestures of fingers, and expressions of the face arose out of Joan Benesh's interest in East Indian Classical Dance. Signs placed above the stave specify these details (Fig. 16).

Soon after the launch of BMN, Joan Benesh met Marianne Balchin, a former member of the Ram Gopal Company. Marianne joined the first graduating class,

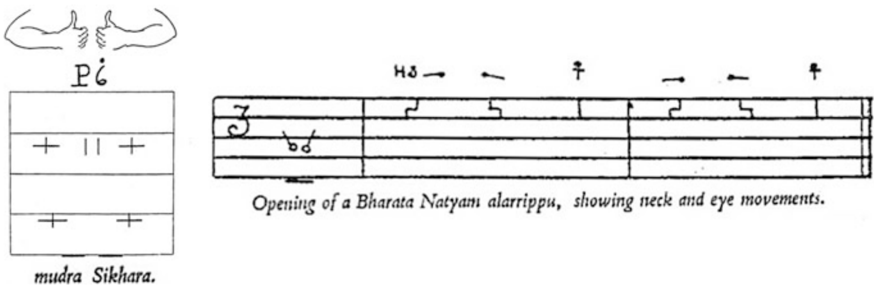


Fig. 16 Two illustrations of Bharata Natyam: finger gestures and eye movements. Notating Indian Dance by Rudolf Benesh, in 1956

which also included students of modern dance, folk and national dance, character dance, historical dance, choreographic analysis, ergonomics and medicine. Rudolf and Joan gave students the responsibility for researching their own practice. The team then worked at developing BMN for recording each style, which led to its extension beyond the realm of classical ballet [10].

3 Specificity of the Benesh System

Watching a film or a cartoon, one forgets that the movement reproduced on the flat screen is made of a succession of snapshots, which are in themselves static. As one starts to read a Benesh score, the movement emerges from the paper as the eyes move from one frame to the other. With the addition of an essential element: the Benesh system analyses the posture in three dimensions in relation to time and space (Fig. 17).

3.1 Several Layers of Information

The organization of the stave allows the reader to abstract body movement and spatial-temporal activity. This enables one to concentrate on one element at a time. For example, as shown in Fig. 18, the reader might focus on the spatial description (below-stave information) but ignore the body position (in-stave information).

3.2 The Information Revealed by Key Frames

Each key frame shows a body position in relation to time and space. The position of the feet provides a sense of gravity, as illustrated in Fig. 19.

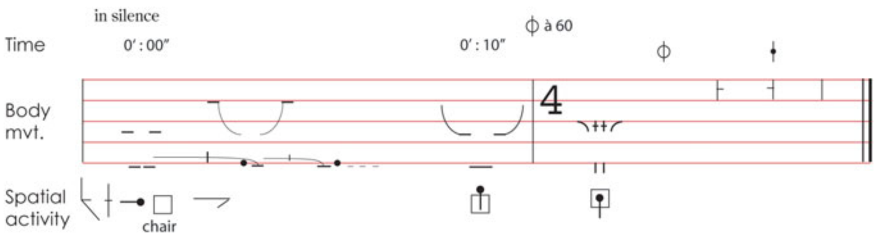


Fig. 17 Organization of the stave

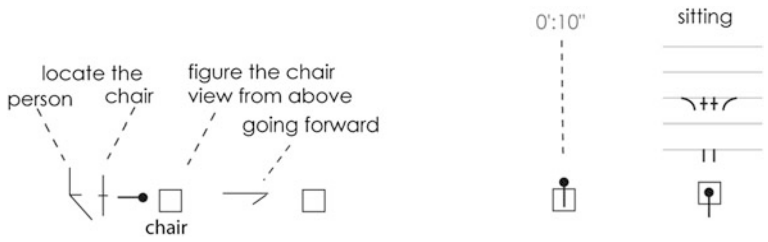


Fig. 18 Starting at the left edge of the room, standing in profile, facing a chair placed in the centre: go forward towards the chair, stop in front of it, facing front; then position yourself on the chair. Standing? Sitting? The information in the stave will tell you

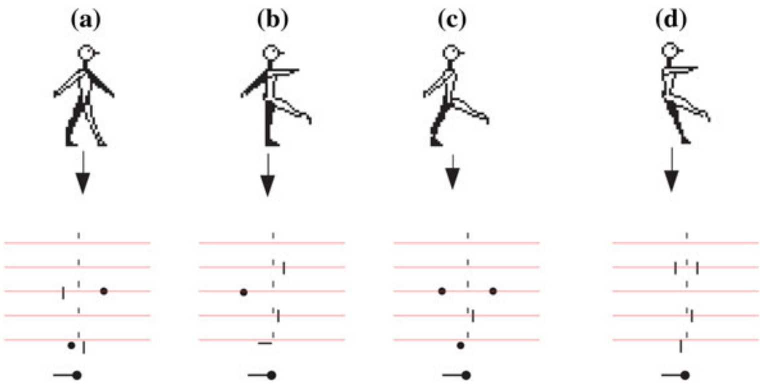


Fig. 19 Four figures viewed in profile, and their notation

Let us concentrate on the notation of the supported feet, remembering the three basic extremity signs illustrated in Fig. 6 (Fig. 20).

This subtlety is particularly useful for the analysis of walking (Fig. 21).

In addition to analysing and abstracting the skeleton in a simple and efficient way, key frames can also record the intention of the movement, as illustrated in Fig. 22.

Fig. 20 **a** Left foot behind, right foot in front: weight distributed evenly on both feet. **b**, **c**, and **d** are on one foot. **b** Standing on left foot level with the body, steady on his leg. **c** and **d**: are both off balance. **c** Standing off-balance on left foot that is behind the body. **d** Standing off-balance on left foot that is in front of the body

Fig. 21 The two first steps end with the weight distributed between two feet (back leg bent, weight on ball of foot); the next two steps end with the weight transferred fully onto the stepping foot (back leg bent, foot lifted off ground)

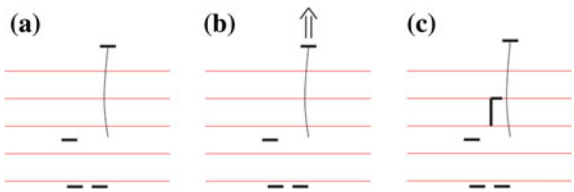


Fig. 22 Right arm lifts directly overhead: **a** a simple arm gesture; **b** the intention is to reach up to the ceiling; and **c** the right shoulder rises as the arm reaches upward. Example **b** specifies the upward movement intention, whereas example **c** shows the skeletal result

BMN can show the body part that leads a movement by using letters, such as A for arms, B for body, etc. Furthermore, adding a number, e.g., B4, may indicate specific body areas (Fig. 23).

The variations of intensity in the execution of a movement can be specified by means of signs borrowed from the vocabulary of music theory (Figs. 24 and 25).

To indicate the state of muscle relaxation or tension, BMN surrounds the *p* or *f* with a circle. This description suggests an inner state, a tonic nuance that may imply emotional expression (Fig. 26).

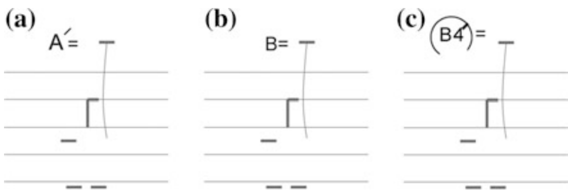


Fig. 23 Three different ways to lead the movement described in Fig. 22: **a** the right arm leads the movement; **b** the body leads the movement; and **c** the right shoulder blade leads the movement

ppp	*pp*	*p*	*f*	*ff*	*fff*
completely relaxed	very soft	soft	strong	very strong	maximum strength

Fig. 24 The vocabulary of music theory

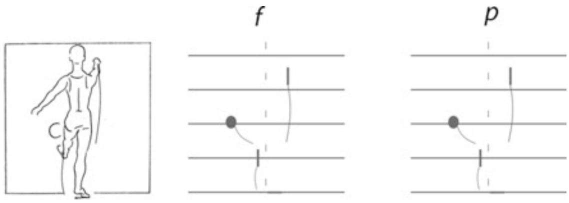


Fig. 25 Two ways to kick a ball, strongly and softly

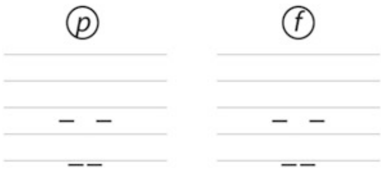


Fig. 26 A relaxed versus a stiff body

3.3 *Movement Lines Visualize Path Direct Versus Indirect*

In the coronal plane—level with the body—the movement lines are the exact reproduction of the path drawn by the hands in space. Figure 27 shows paths of the hands opening outward then upward (arms remain straight) while Fig. 28 shows paths of the hands lifting upward and outward (arms flex then extend).

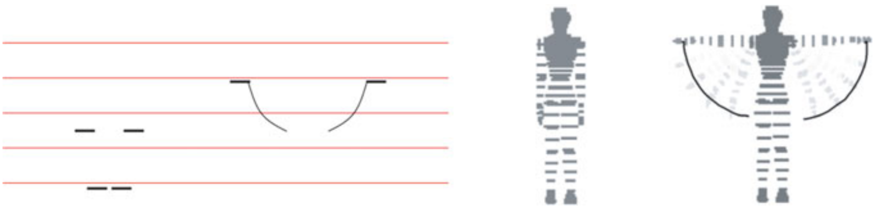


Fig. 27 Lifting both arms in extension to the side, away from the body, up to just below shoulder height

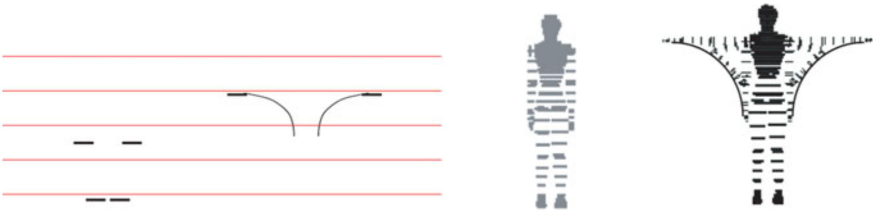


Fig. 28 In the *left* example, movement lines summarize the folding/unfolding action

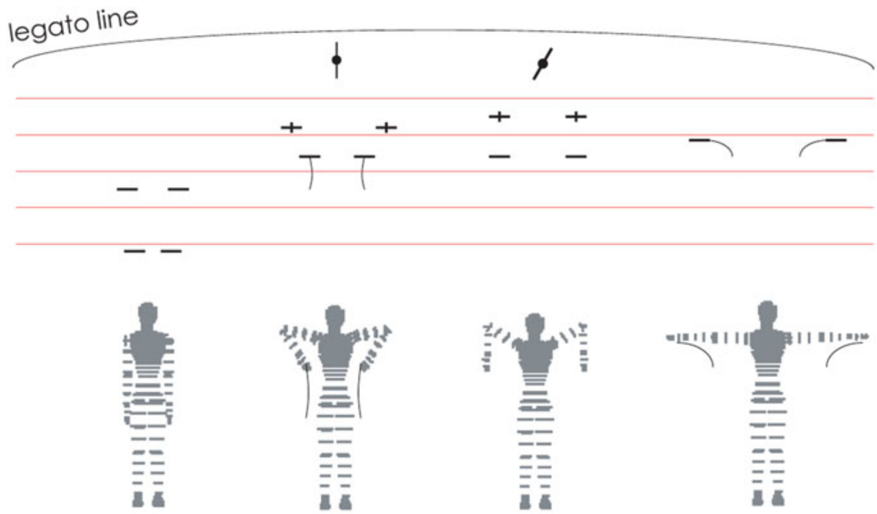


Fig. 29 Break down of the movement paths

According to the degree of accuracy required, key frames will break down the movement paths as in Fig. 29. An addition of rhythm signs will maintain the timing of the movement. A legato line will phrase it and render the fluidity of the movement.

The following examples show movement in the transverse plane (Fig. 30).

Fig. 30 The hands start in front at shoulder height. In the *left* example, movement lines summarize the horizontal path of the hands as they move outward then backward. In the *right* example, key frames break down the movement paths

4 BMN and Technology

4.1 Software

4.1.1 From ChoreoScribe to MacBenesh

ChoreoScribe, developed at the University of Waterloo (UW) Computer Graphics Laboratory in the early 1980s, was the first main frame computer software for creating and editing BMN [11, 12]. Continuing from this project, MacBenesh, the first personal computer software, emerged in 1984 from collaboration between UW and the Ontario Science Centre. The notation consultants were Professor Rhonda Ryman, an expert in Benesh and Laban notation systems, and Robyn Hughes Ryman, Choreologist with the National Ballet of Canada [13].

MacBenesh, an editor for single-dancer BMN scores, was developed on the Macintosh computer, which offered enhanced graphics capabilities and enabled advanced end-user interface features. It was further developed and commercialized by DanceWrite, until development ceased in 1992. Consequently MacBenesh runs only under Mac OS9. MacBenesh proved helpful to this author in writing the professional course for the Conservatoire de Paris when, with the transition to Mac OS10, she succeeded in importing all the defined Benesh signs into a vector graphics editor.

4.1.2 Benesh Notation Editor

The Benesh Notation Editor (BNE) is a PC Windows software program for creating and editing multi-dancer BMN scores, developed by the Benesh Institute in collaboration with the University of Surrey. It acts as a 'word processor' for the notation, enabling the production of quality multi-stave printed scores that can be edited, copied, stored digitally, printed and transmitted by e-mail just like any other file [14].

The BNE was developed to meet the needs of professional notators, but it may also be useful for notation teachers and students, who may find that they can produce quality scores more easily and quickly than writing them by hand.

4.2 BMN and DanceForms

DanceForms is choreography software developed by Vancouver, B.C. based Credo Interactive, Inc. Given that Rudolf Benesh visualized movement as a series of frames, there is an obvious similarity between BMN and DanceForms. Both notator and animator aim to capture a succession of key frames representing body positions over time [15] (Fig. 31).

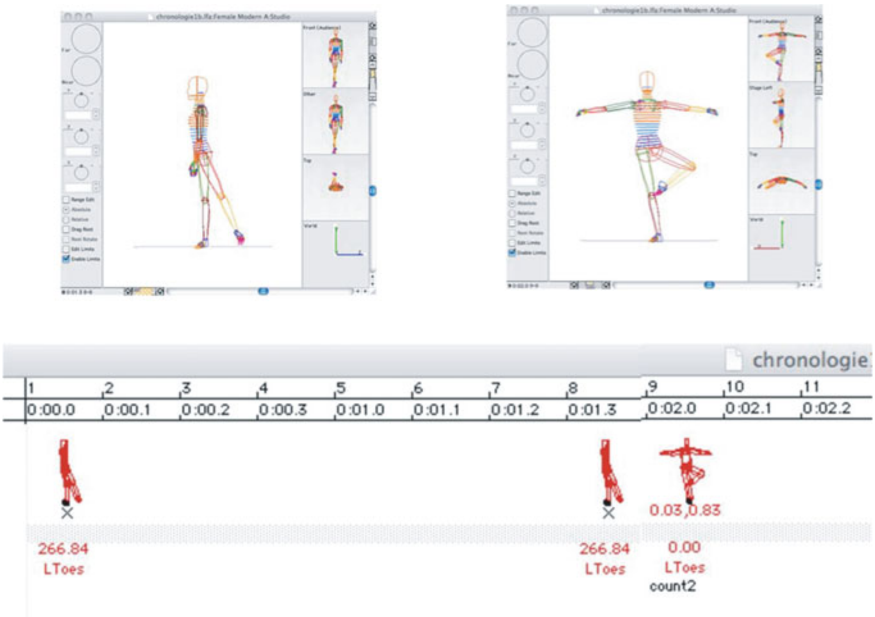


Fig. 31 Screen capture of a DanceForms animation The *top* illustrations show the Studio window for two body positions, created with Ballet Moves palettes. The *lower image* shows key frames placed in sequence in the score timeline

The resulting product is, however, quite different. In notation, the body is abstracted and it is up to the reader to reconstruct the movement as the eyes move along the score. By contrast, computer animation uses various modes to represent the body, and it is up to the viewer to decrypt the organization of the movement.

Animated examples can clarify notation theory without using words, which are often distorted and misunderstood. Beginner students often have difficulty understanding how the three layers—time, body movement, and spatial orientation—are interconnected. For example, altering the placement of a direction sign or a turn sign will change the resulting movement (Fig. 32).

The emergence of e-books that blend graphics, writing, images, and movies encourage the use of these medium. The recent iBook, *Benesh for Ballet, Book 1*, published by Rhonda Ryman and Robyn Hughes Ryman, can only be praised as an example, combining notation, animation, and word description.

5 Adapting BMN for Special Contexts

"Notation is a tool, and it is not an end to itself." (Rudolf and Joan Benesh 1956, p. 5) To serve a purpose, the system must be flexible for better efficiency.

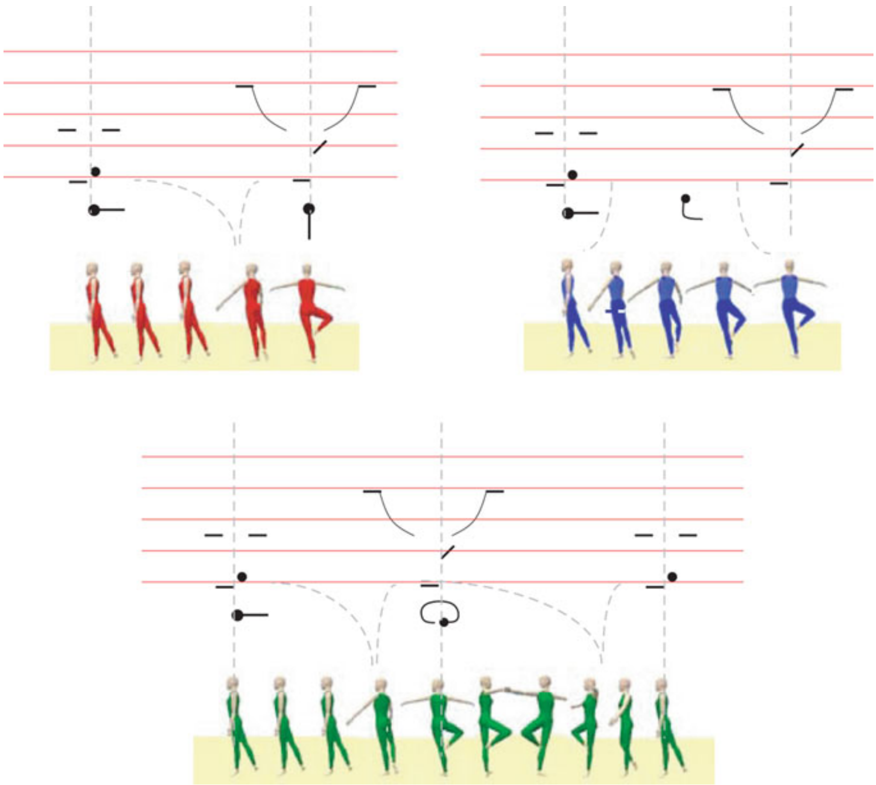


Fig. 32 The animations clarify the concept of direction/turn sign placement

In 1959, Rudolf contributed to a project initiated by the Paris Centre of Technical Studies for the Clothing Industry. He observed and notated the various manipulative skills of dressmakers and the way they used their machines. He observed and notated the way dressmakers worked at their machines and manipulated their fabrics. His aim was to record relevant information only. To focus on the upper body and arms, he adapted the five-line stave as illustrated in Fig. 33 (Unpublished notes. Benesh Institute, London, 1959).

Since the dressmaker's arm movements occurred only in front of the body, there was no need to distinguish hands in front of, level with, or behind the body. To plot the hands on this newly defined stave, it was possible to use only wrist direction signs, normally added to indicate arm rotation. To plot the hands at precise widths, he added location signs, normally used to locate individuals in the working area, as illustrated in Fig. 34 for a recent project.

In 1976, BMN was adapted to study the movement and posture of sitting subjects [16]. For this project, Rudolf adapted a five-line stave for the chair, as illustrated in Fig. 35: The top half of the stave represents the back of the chair, and the bottom half of the stave represents the seat. The double line on the stave indicates

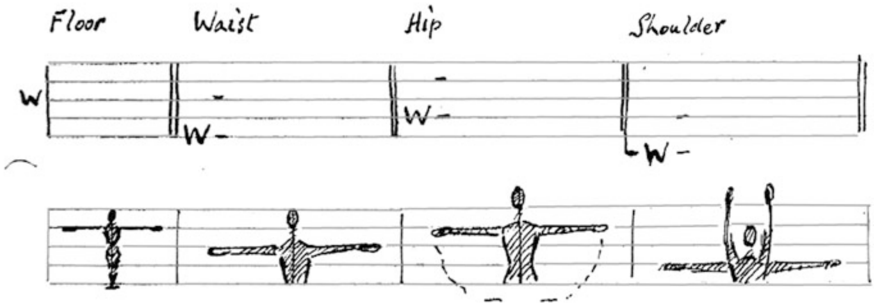


Fig. 33 To focus on the upper body and arms, Rudolf Benesh adapted the five-line stave by "zooming in" and redefining the anatomical landmarks represented by stave lines

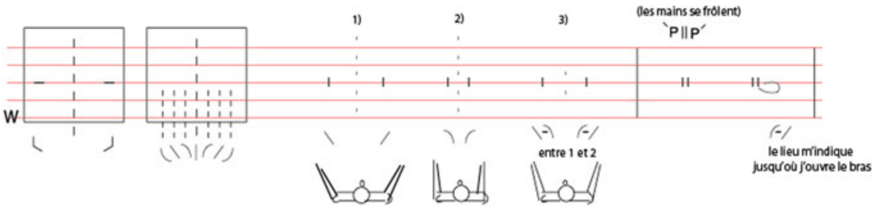


Fig. 34 A digital application of Rudolf Benesh's unpublished notes

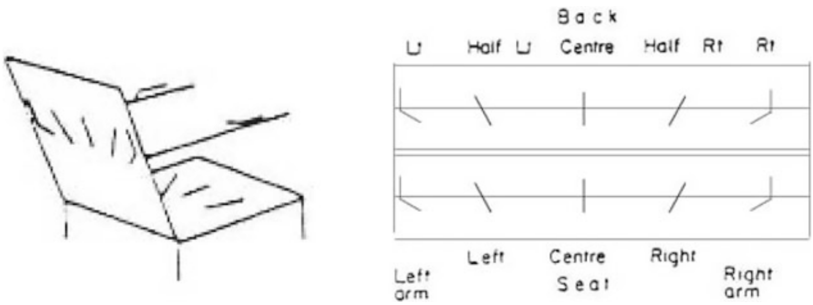


Fig. 35 The adapted stave and locations signs used to define parts of the chair

where back of the chair and the seat intersect. The chair is viewed as if from behind, i.e., the left half of the frame represents the left side of the sitter. To signify which part of the chair is providing support, the standard location signs are used (Fig. 36).

The above examples show the adaptability of BMN. I was able to build on this flexibility in the following two cases.

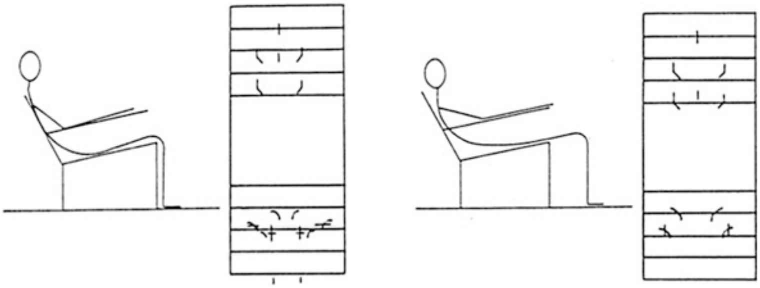


Fig. 36 The chair stave and below the posture analysis

In the first case, I was asked to assist a PhD student whose thesis topic was musculoskeletal disorders of the thoraco-lumbar region among grave diggers. He needed a way to clarify specific elements in his research photographs [17] (Fig. 37).

In the second case, I collaborated on a research program in the field of Human Computer Interaction. In this case, I applied the adapted stave Rudolf Benesh developed for the 1959 study quoted above (Unpublished notes. Benesh Institute, London, 1959).

The basic principles of BMN allow application by expert notators to a wide range of research problems, provided that each specific adaptation or extension is documented in context.

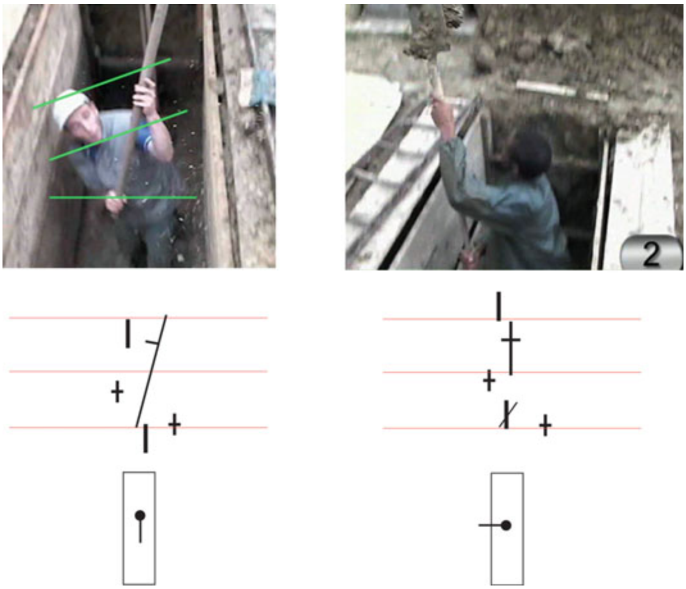


Fig. 37 BMN provides a way of recording each subject's posture from the photos, allowing comparison and analysis of the differences

6 Conclusion

This paper gives the rudiment of Benesh Movement Notation, which has proved its efficiency in a variety of applications. Considering its application to the robotic field, BMN answers the need to identify key frames on a time line, to show the path of movement and to locate the body in space. It analyses the body scheme and focus on dynamics of movement as well.

To apply BMN to the robotic field, it requires a deep knowledge and that take years. Within the framework of humanoid robot motion, the use of movement notation makes sense in collaboration between the notation expert and the scientist.

Acknowledgements I am extremely grateful to all my Benesh colleagues who left few but essential traces of the genesis and the philosophy of the system. In particular Joan Benesh who drawn up much of her husband unpublished writings in one book which remain a major testimony [11]. Julia McGuinnes-Scott who introduced BMN in medicine, and anthropology in particular [1], Marguerite Causley, who exposed BMN in physical education [18]. My recognition to the Benesh Institute, and in particular to Liz Cunliffe who gave me access to Rudolf Benesh's notes and to Adrian Grater who commented them to me.

My thanks also to Rhonda Ryman and Robyn Hughes Ryman, for their proofreading and expertise to the English expression towards the Benesh notation.

References

1. J. McGuiness-Scott, *Movement Study and Benesh Movement Notation* (Oxford University Press, Oxford, 1983)
2. K. Wolf, C. Le grand, X.Y. Compagnie, Benesh score, Hector Berlioz Library of the Conservatoire de Paris (2011)
3. K. Wolf, Écrire le cirque? La notation du mouvement Benesh pour les arts du cirque, In memento3, Hors les murs (2011)
4. R.J. Neagle, K. Ng, R.A. Ruddle, Developing a Virtual Ballet Dancer to Visualise Choreography, In *Proceeding of AISB Convention*, March–April 2004
5. R.J. Neagle, Emotion by Motion: Expression Simulation in Virtual Ballet, Ph.D. thesis, University of Leeds, 2005
6. G. Politis, *A Translator For Human Movement Notation* (University of Sydney, 1989)
7. G. Politis, D. Herbison-Evans, Computer Choreology Project at the University of Sydney. Leonardo **21**(1), 34–38 (1988)
8. S. Saad, D. De Beul, S. Mahmoudi, P. Manneback, An Ontology for video human movement representation based on Benesh notation, in *2012 International Conference on Multimedia Computing and Systems (ICMCS)*, May 2012, pp. 77–82
9. E. Mirzabekiantz, *Grammaire de la notation Benesh* (Centre National de la Danse, Paris, 2000)
10. R. Benesh, J. Benesh, *Reading Dance: The Birth of Choreology* (Souvenir Press, London, 1977)
11. D. Dransch, J. Beatty, R. Ryman, *ChoreoScribe: A Graphics Editor to describe Body Position and Movement Using Benesh Movement Notation* (University of Waterloo, 1986)
12. R. Ryman, B. Singh, J. Beatty, K. Booth, A computerized editor for Benesh movement notation. CORD Dance Res. J. **16**(1), 27–34 (1984)

13. R. Ryman, R. Hughes-Ryman, The MacBenesh editor: a “Word Processor” for Benesh notation. Dance Notation J. **4**(2), 16–26 (1986)
14. Royal Academy of Dance Homepage, The Benesh Notation Editor. http://www.rad.org.uk/study/Benesh/benesh-notation-editor. 2015
15. Credo Interactive Homepage, Credo Interactive Inc. DanceForms 2.1. http://charactermotion.com/products/danceforms/. 2015
16. P.A. Kember, The Benesh Movement Notation used to study sitting behaviour. Elsevier Appl. Ergon. **7**(3), 133–136 (1976)
17. P. Simonet, L’hypo-socialisation du movement: prevention durable des troubles musculo-squelettiques chez des fossoyeurs municipaux, Ph.D. thesis, Conservatoire national des arts et métiers, 2011
18. M. Causley, *An Introduction to Benesh Movement Notation: Its general Principles and its Use in Physical Education* (Max Parrish, London, 1967)

The Origin of Dance: Evolutionary Significance on Ritualized Movements of Animals

Satoshi Oota

Abstract Dance is not a human-specific behavior. It is well known that some non-human animals perform “dances” for various purposes, some of which are related to food acquisition and courtship. According to the traditional Darwin’s evolutionary paradigm, our genetic information is subject to natural selections, or more boldly speaking, a principle of “survival of the fittest.” In these contexts, dance or ritualized movements performed by animals should have clear evolutionary traits. For example, some birds perform ritual dances to draw attentions of the opposite sex to mate with. Under competitive circumstances, the “dance” should play a critical role for individuals to propagate their own genetic information in the population. However, recent studies also suggest that human has genes associated with creative dance performances, which are not directly relevant to the Darwinian selection. I will briefly describe behavioral characteristics of non-human “choreographed dances” and discuss their evolutionary significance, as well as similarity and difference between the non-human and human neurobehavioral characteristics.

1 The Origin of Dance

Human is often conceived as a unique organism on the earth: i.e., we have highly developed intelligence, consciousness, and languages, which are supposed to be incomparable with the other non-human species. Some of readers of this book may have never doubted the anthropocentrism, or the superiority of human over the other species.

However, biologists typically have a different perspective on the human: i.e., *Homo sapience* (modern human) is one of countless species on the Earth. In other words, *H. sapience* is just one of leaves of a huge phylogenetic tree of the Kingdom Animalia (Metazoa).

S. Oota (✉)
RIKEN Bioresource Center, Tsukuba, Japan
e-mail: oota@riken.jp

J.-P. Laumond and N. Abe (eds.), *Dance Notations and Robot Motion*,
Springer Tracts in Advanced Robotics 111,
DOI 10.1007/978-3-319-25739-6_15

One good idea to elucidate a phylogenetic position of human is to compare us with chimpanzee, *Pan troglodytes*, which is our nearest neighbor. We share approximately 94–99 % (depending on a measure) of genetic information and many phenotypic traits [1], including "verbal" communications [2], symbolic competency, complex multifaceted problem-solving capabilities [3], and social interactions (namely, politics [4]).

But still you might think that they, "apes," are inferior species, and human is an aloof species from the others. It is true that there is only one kind of "human" on the Earth [5]. In other words, there are no subspecies of human, unlike the other many animals. Because of this fact, it may look reasonable to regard human as a "special" and "exceptional" organism. And you may think that it is trivial and self-evident. However, as a matter of fact, there used to be several "humans" on the Earth: at least three kinds of humans (*Homo heidelbergensis*, *Homo neanderthalensis,* and *Homo sapiens)* coexisted approximately 50,000 years ago [6, 7].

Today the genome data of the extinct *H. neanderthalensis* (Neanderthals) are available [8, 9]. According to the comparative genome analysis, there is no reason to refute an assumption that they almost behaved like *H. sapiens* (modern human) especially regarding intelligence, consciousness, and probably also creativity. In fact, many pigment blocks were found at European Mousterian sites, suggesting that Neanderthals decorated their bodies with them [10]. And it is possible to speculate that they performed ritualized movements in their social context [11]: i.e., they might perform choreographed dances [12].

This kind of non-anthropocentric approach about human behaviors immediately raises another question: how general are the ritualized behaviors across non-human species? Surprisingly, the ritualized behaviors, or dance-like movements, are observed in many non-human animals [13]. Even some invertebrates are known to reveal ritualized movement patterns [14, 15].

2 Dancer in the Dark

The most famous "dance" performed by an invertebrate is honeybee's waggle dance [16, 17]. In 1946, Nobel laureate Karl von Frisch made a brilliant discovery that honeybees waggle to tell the other bees a polar coordinate of patches of flowers (a direction and a distance from the hive) with respect to the sun as a reference point [18, 19]. This amazing discovery fascinated many researchers: how can their simple neural circuitry handle this complicated nonverbal communication between bees?

Many researchers have been tackling this problem in various paradigms: neuroscience [20], ethology [21], linguistics [22], artificial intelligence [23], and robotics [24]. And researchers gradually began to understand how intriguing this problem was.

For example, we can perceive honeybee's dance because we are humans, who are able to look over the hive. How can honeybees "recognize" and reconstruct their own dance patterns especially in the dark hive? A new research suggests that bees track trails of olfactory compounds (semiochemicals), which they emit during the dance [16].

This discovery about non-visual perception of honeybees brought a new riddle: it is undoubted that they use spatiotemporal patterns to recognize their information. In this regard, how can they code and decode the symbolic communication made by the olfactory trails? The spatial resolution and the effective range of semiochemicals should be quite indefinite for bees jostled away from a dancer. After all, even this kind of "basal" issue is still an open question.

Some research groups are tackling the honeybee dance with non-biological approaches: a robotics group is trying to mimic the honeybee's waggle dance in an algorithmic way [24]. These kinds of top down approaches began to elucidate the complicated bee's nonverbal communication in a different perspective: i.e., as a collective behavior [25]. A Japanese group elaborated on errors of the bee's navigation system, by modeling bee's behaviors with a Markov process [26]. Their study shows how bees decide their behaviors to reach a goal in the natural foraging environment, which can rapidly and dynamically change.

But those studies mainly focus on the foraging and foragers, not the bee's dance itself. So far, the neuromotor control of the intriguing waggle dance has been treated as a black box. How did honeybees evolve the neural circuitry to control the complicated choreographed dance? One idea to tackle this problem is to segment the complicated waggle dance into simple elements. Even the complicated dance is composed of simple and common motions [27], which are potentially shared by the other species, like fruit fly (Drosophila: one of the representative genetic tools). This traditional approach may shed light on the riddles of the waggle dance.

3 Darwin's Cruel Dance

Crayfish displays complex but obviously ritualized movement patterns when two males engaged in pseudo-copulatory behavior [13]. This characteristic behavior plays a significant role to suppress extreme aggression between the two male opponents, one of which would be potentially slaughtered by the victor if the ritual motion were not performed. The dominant-subordinate relationships are well analyzed in terms of survival rate of the subordinate after their aggressive encounters. It is noteworthy that female is always in a subordinate position (due to her physically inferior condition) and she can be killed if she refuses the male's attempts to mate.

Such observation is consistent with Darwin's paradigm that fitness (in terms of population genetics) is to be optimized through their "dance:" it is obvious that if victors (dominants) massacre majority of subordinates, their gross fitness will considerably decrease in the next generation.

The pseudo-copulation performed by two males has no physiological meaning at all, but it probably contributes to maintain the species, as well as its genetic diversity of the crayfish population. In an evolutionary context, it is quite interesting how crayfish acquired such hard-wired neural circuitry to perform the "unnatural" behavior.

In terms of Darwinism, it is possible to hypothesize that subordinate males that perform pseudo-copulation can survive (raise their fitness) and the other subordinate variations just almost perished. While this hypothesis seems plausible, we still need very strong assumptions: (1) there were sufficient variations in the ancestral crayfish population in terms of pseudo-copulatory behaviors; (2) presumably small number of (or single) mutations are determinants of the variation: i.e., the neural circuitry that control the pseudo-copulation mutations is a result of relatively minor mutations in an ancestral form.

This implicates that a basal neural circuitry to perform copulation (namely, sexually heterogeneous copulation) is presumably shared by female and male and the small number of mutations devised a new mechanism that can switch mail-type neural circuitry to the female-type only when the individual happens to be subordinate to an aggressive victor (otherwise, the subordinate cannot normally mate females).

An important issue here is that evolution cannot suddenly design a complicated biological functions (at least according to the modern evolutionary theory) [28, 29]. If organisms happen to acquire some functionality, its prototype must have been prepared in advance. In the crayfish case, it is reasonable to assume that the small number of mutations made the copulation-related behavior diverged between female and male [30], and another small number of mutations reverted the male copulation-related phenotype to the female type under certain conditions. This explanation looks roundabout, but biologically plausible.

Today it is possible to computationally "simulate" the neuromotor mechanism in silico [31]. This kind of top down approach will be useful to analyze the pseudo-copulatory behavior under controlled conditions.

The evolutionary process is more conservative in vertebrates rather than invertebrates due to complexity of the biological systems: e.g., vertebrates share more neuromotor systems than invertebrates [32]. In this regard, the neural circuitry relevant to ritualized (dance-like) motions is also expected to be more evolutionarily conserved within vertebrate species than within invertebrates.

4 Bird Dance on a Chorus Line

The ritualized behaviors of honeybees and crayfishes are good examples that whole body-level movements can play important roles at intra-species interactions, presumably at the population level. But we should note that, at least, the crayfish's "dance" behaviors are purely innate and have no flexibility based on new experiences or acquired memories: i.e., a choreography of the Darwin's cruel dance is thoroughly determined only by the genetic information.

It is well known that birds literally sing in their courtships [33]: in our context, the "song" implicates that they chirp with a nonrandom structured pattern, which is retrieved from and/or generated with elastic memory: i.e., imitation rather than instinct. So their songs are worth to be genuinely called "music." Recent studies suggests that birds are outstandingly "intelligent" organisms [34], but in a slightly different way from mammals, including apes [35].

Likewise, there are many reports about avian ritualized movements, which are extremely complicated and aesthetic [36–43]: at least for human eyes, their movements look genuine dances with elaborate songs or music. For example, zebra finches (*Taeniopygia guttata*) reveal dances composed of beak movements, head motions and hops, which are associated with song [44]. According to the study, the correlation between the dance movements within fathers' and son's songs was surprisingly significant: i.e., the choreography of the dance patterns is probably transmitted from tutor to pupil together with the song.

The other outstanding (and almost breathtaking) avian dances are beautifully filmed in a BBC documentary, "Planet Earth: Bird of paradise [45]."

Birds have an excellent visual perception [46]. Many mammals, like rodents, largely rely on an olfactory perception to live, probably because their ancestors were mainly nocturnal when they were threatened by predators like fierce dinosaurs. On the other hand, birds, descendants of the extinct dinosaurs, need long-range sensing capabilities with electromagnetic radiation (i.e., visible light) to fly, presumably by which they evolved the excellent visual perception. It is possible to speculate that their highly-developed visual perception in turn provides them unique courtship behaviors.

First of all, many birds are outstandingly beautiful. This characteristic has two-fold implications in terms of Darwinism: (1) more visually conspicuous individuals have a tendency to have higher fitness; (2) they mainly rely on visual information to choose their mates (usually for a female to choose a male), not a practical capability to survive and/or reproduce. In other words, if a bird is aesthetically more dominant, the individual will have higher fitness.

This looks almost irrational in terms of Darwinism. But regarding the ritual movements, i.e., dance, their aesthetics is more explainable as follows.

Our genetic information is fragile. Radiation (mainly cosmic rays) [47] and some chemicals are always bombing our bodies at molecular level [48]. As a result, the bombardments cause countless mutations in our genome. The rough estimation of the mutation rate of mammals is the order of 10^{-9} per site per year in the germ line (genetic information transmittable to the next generation) [49, 50]. Considering that our genome size is 3234.83 Mega bases (the order of 10^9; NCBI genome ID: 51), the mutation rate is high enough to continuously cause some genetic abnormalities. In fact, many of our diseases are caused by mutations: e.g., neurological disorders [51], psychiatric disorders [52], and cancers [53].

The ritualized movements of animals are typically standardized or fixed: they are like an audition of stage actors/actresses before a stage director. Namely, their dances are very strict screening tests to examine their motor functions. If some mutations impair the motor functions, they will be easily detected as a behavioral phenotype during the audition. Note that neuromotor functions are extraordinarily complicated in terms neural circuitry and the musculoskeletal system, and many genes (units of genetic information in the modern sense) are involved in normal activities of the system. If a single mutation impairs a long cascade of the systematic dependencies, the whole system will potentially collapse.

Therefore, the ritualized movements can be regarded as a very strict and cruel screening test to exclude potential genetic abnormalities, which may lower the gross fitness of the population.

In this context, the aesthetics on appearance of birds may also be explained. It is well known that plumage colors are linked with some genes [54–56]. For example, in White Plymouth Rock and its F2 generation with white plumage, an insertion of an avian retroviral sequence in the tyrosinase gene is reported [57]. In this case, the plumage color can be a good marker to detect the mutation around the tyrosinase gene.

Thus, a female can choose a healthier male (with less mutations) based on the plumage colors. Of course, it is also possible that a mutation gives a positive effect to an individual in terms of conspicuousness: e.g., albinos. However, this is another issue.

For many birds, due to their extraordinary visual perception, "dance" may occupy a special position in ethology, especially in terms of courtship displays [58–60].

5 Almost Human: Chimpanzee

Recently researcher found that our closet relative, chimpanzee, has very complicated and human-like social behaviors, rather than we thought before [61]. For example, for long time it has been believed that "warfare" is a human-specific behavior. However, it is reported that a group of chimpanzees attacked against the other group of chimpanzees in an organized way [62–64]. It is also reported that (typically male) chimpanzees' aggressive behaviors contain massacres, infanticides, and rapes [61, 65–67]. What they did to the same species looked very like human violence, some of which combatants would potentially do to their enemies and/or noncombatants in wartime [68, 69].

We should note that ants and bees in fact occasionally show mass assaults on their subspecies: e.g., hornets attack against honeybees to plunder larvae as foods [70]. While they look "wars" for human eyes, ants and bees are actually regarded as "super-organisms," which behave as if the groups were individuals. Therefore, we should call it "duel" rather than "war." Besides that, they rarely attack on the same species, as well as never kill just for fun (in the sense of human-like "sadism").

Chimpanzees, surprisingly, kill just for fun [67]. Considering such observation about the chimpanzees' decadence, it is not so surprising that they reveal (human-like) cultural tastes. Unfortunately, except for large-scale massacres, it is not easy to observe their "civilized" aspects in the harsh and competitive wild environment.

In an artificial environment, Hattori et al. reported that a chimpanzee revealed synchronized tapping to an auditory rhythm [71]. One might criticize their study that the chimpanzee just mimics human's complicated motion patterns, which are occasionally observed in rearing environments. But, according to Hattori et al., the chimpanzee spontaneously (presumably by its own will) started the "tap dance" with the sound of music after certain training.

If it is genuinely true, we can tell that a chimpanzee can "play" or at least "track" rhythms with spontaneous motor functions. These behaviors are pretty close to "dance" in an ordinary sense.

6 Another Human: Neanderthals

Of course, human dance is not mere tracking of rhythms. Human often dances to express internal emotions: human dance is one of the nonverbal communications.

The nonverbal communications greatly play important roles in our social contexts. First of all, in conversation, we fetch a great amount of information from facial expressions, which generated by facial muscles innervated by the craniofacial nervous system. Without it, our communication capability would be greatly limited.

Secondly, our "gesture" also provides significance to modify the verbal expressions: e.g., head- and body-movement plays significant roles in psychotherapy [72], suggesting that even subtle movements carry significant emotional signals. Human gestures, postures, and movements also affect mate selection [73], which are regarded as prototypes of dance in the context of courtship. This issue directly associates dance with evolution in terms of sexual selection [74]. But still our verbal communication plays the most critical roles for the communication, which greatly influences the human culture.

Here we can cast a simple but insightful question: why did the Homo subspecies go extinct except for us?

According to genetics studies on human history, our effective population size (repeatedly) experienced bottlenecks [75–78]: i.e., we had been nearly wiped out from the Earth. While the time and the period of the bottlenecks are still controversial, a cold period approximately 39,000 years ago called Heinrich Event 4 (H4) is thought to be one of our crises [79, 80]. Of course, the harsh environment also affected the other human subspecies. Regarding the extinction of Neanderthals, there are mainly two competing hypotheses: (1) Neanderthals were unable to adapt to the rapid climate change [81]; (2) Their extinction was rather due to competition from anatomically modern humans (AMH) [82]. Recent studies suggest that the latter theory is more likely.

Then what kind of traits of AMH brought advantage over their competitors? Some researchers speculate that AMH could surpass them in sophisticated social relationships or cooperation [83]: i.e., we have had an advantage over the other subspecies in terms of communication capabilities, especially motor functions associated with the articulate speech [84].

This speculation is based on comparative analysis on Forkhead box protein P2 gene (*FOXP2*) [85], which is also called the "language gene."

In 1990, Hurst and colleagues studied a three-generation pedigree, known as the KE family, some of whose members suffered from a developmental verbal dyspraxia. From the genetic analysis, in 2001, Fisher and colleagues finally identified *FOXP2* as a gene responsible for the speech and language disorder.

Interestingly, some songbirds have *FoxP2* very similar to human *FOXP2*. For example, zebra finch *FoxP2* is approximately 98 % the same as the human homologue, which makes vocal learning possible.

In 2009, S. Pääbo's group "humanized" the mouse *Foxp2* allele by introducing two amino acid substitutions, which seemingly affects the mouse basal ganglia: i.e., medium spiny neurons increased dendrite lengths, as well synaptic plasticity [86].

H. Sapience and *H. neanderthalensis* (Neanderthals) share the two mutations (substitutions) in the *FOXP2* gene, which presumably gave us speech capabilities. The haplotype was subject to a strong selective sweep: i.e., the human and Neanderthal *FOXP2* s are positively selected, suggesting that the articulate speech capabilities significantly raised our fitness.

Meanwhile, in a non-coding region, S. Pääbo's group (again) found a human-specific substitution in a regulatory element, *POU3F2*, which controls an expression level of *FOXP2* [87]. This supports the above speculation: i.e., modern humans (precisely speaking, anatomically modern humans (AMH)) may have dominated the survival competition over taciturn Neanderthals owing to the articulate speech capabilities.

As I mentioned, dance is a tool for nonverbal communication. And probably Neanderthals sang and danced [88]. Then what kind of dance did they perform? Did Neanderthals choreograph an elaborate dance to compensate their inferior verbal communications?

At this point, the natural science is too powerless to elucidate their lost performances. We can just imagine their behaviors via fictional works, like "Dance of the Tiger [89]." But they surely existed and silently left the Earth.

7 Dance as Aberrant Motions

Chorea ('dance' in Latin) refers to any kinds of neurological diseases marked by involuntary and/or convulsive movements, especially of limbs. Of course, despite of the taste of this Latin term, this word "chorea" has no direct implication associated with actual dance. Typical chorea is a genetic disorder caused by a single mutation in a gene coding the huntingtin protein (the gene is also called the huntingtin or *HTT*). This chorea is called Huntington's disease (HD) after the comprehensive description by George Huntington in 1872.

While HD is a relatively rare disease in the general population (only 5–10 cases in 100,000 persons), this disease is known as a lethal condition: death typically occurs within 15–20 years of showing the first signs of HD. The patients suffer from severe progressive symptoms, including mental disability.

HD drew much attention due to its characteristic genetic traits. Many movement disorders are associated with multiple genes. This means that we need to consider complicated (and quantitative) interactions between them in terms of epistasis and/or quantitative genetics [90, 91]. Meanwhile, HD has only one causal gene: *HTT*. Furthermore, the huntingtin protein has a kind of discrete indicator that determines degree of severity of the diseases: the number of CAG (cytosine-adenine-guanine) triplet repeats, which code a polyglutamine tract ("polyQ"). This mutant protein is putatively a toxic substance that damages nervous cells.

It is noteworthy to mention why the HD is called chorea. The disease has been already known since the Middle Ages [92]. People recognized this disease by certain aberrant motions. But what is the definition of the "aberrant" motion? It is actually not a trivial question. We just know or feel what is the aberrant motion almost instinctively (probably because such capability is crucial for our survival [93]).

Of course, chorea is an extreme example. In our daily life, we occasionally encounter "unusual" behaviors [94]. For example, we can "sense" unusual emotions of our friend through her/his very subtle movements deviated from a known "normal" pattern [95]. We cannot clearly explain the reason. But we can recognize it before we actually know what is wrong with the friend: i.e., we have a capability of empathy for others.

The deviation from "normal" motions well explains what is the dance in the social context. If you are in a competitive environment to acquire your mating partners, first of all, you have to be perceived by your potential partners. In other words, you must not be buried among the others. In the evolutionary context, the deviation can be your advantage over the others.

Of course, chorea is involuntary movements, as well as an instinctive "mouse dance," which are cause by genetic abnormality [96]. On the other hand, authentic dances are voluntarily performed by healthy performers. What I exemplified here is essentially different from ordinary choreographed dances.

Biologically speaking, however, certain involuntary and voluntary movements sometimes share a basal mechanism: i.e., the both movements can be governed by overlapped neural commands [97, 98]. It is hard to refute a claim that certain spectra (the complexity and ranges) of motor commands and the geometry of the musculoskeletal system are, no matter which they are involved in involuntary or voluntary movements, primarily determined by the same genetic information.

According to the paradigm of geneticism, we are human just because we have the human genome. Our neural circuitry and muscle strength, joint stiffness, joint angle ranges are all determined by our genetic information. For example, we don't have as much strength as grate apes have [99]. We cannot perceive four kinds of colors as some species can [100] (except for anomalous trichromacy due to genetic aberration [101]). We cannot fly. We can hardly hear ultra sound [102]. Because we have the genome of *H. sapience* and lack such responsible genes.

Setting an evaluation on the Genetics fundamentalism aside, it is still plausible to assume that basal elements of motor command patterns to govern complicated (dance-like) movement are already embedded in our genetic information. In other words, the elemental motor commands are "hard-wired" and independent from the complex somatosensory feedback. It is also possible to assert that aberrant but voluntary movements are realized by intentionally "blocking" or "alter" certain pathways of the evolutionarily designed neural circuitry, basal part of which is the "hard-wired" architect directly determined by our genome. In fact, it is possible to mimic aberrant motions by a healthy subject.

While such assertion looks rather trivial, this implies an important concept about the origin of complex motions: for example, as I mentioned above, complex bee's waggle dance can be divided to simple and common motion patterns, each of which flies can potentially perform. Obviously, each of their motions is genetically deterministic, but the combination of the elemental motions is exogenic. Therefore, it is plausible to state that even our voluntary movements (e.g., dance) may genuinely share a basal mechanism with bees' waggle dance, which is primarily instinctive but dynamically changeable corresponding to the external environments.

One example of the basal mechanisms shared by vertebrates and invertebrates is the central pattern generator (CPG). Evolutionarily speaking, studying on lamprey CPG is a great interest because this organism is phylogenetically positioned between vertebrates and invertebrates.

To truly understand human aesthetic motions (like beautiful dance), ironically, we may need to elaborate upon abnormal motion patterns caused by some neurological diseases, like chorea, as well as genetically determined (intrinsic) motion patterns of animals, like the honeybee's waggle dance.

8 Dancing Genes (or Dance as a Phenotype)

In 2005, Bachner-Melman and his colleagues performed an interesting analysis on genes that might be associated with the human dancing "phenotype" [103]: they genotyped 85 dancers (3 males, 82 females) and their parents for the serotonin transporter gene (SLC6A4) and the arginine vasopressin receptor 1a gene (AVPR1a). They also genotyped elite athletes and a group of nondancers/nonathletes as control groups.

The serotonin and the arginine are representative neurotransmitters, which putatively play important roles in our mental and physical activities. According to a behavioral neuroscience research [104], for example, serotonin contributes to spiritual experience and empathy [105]. Arginine vasopressin plays an important role in courtship behavior of zebra finches [106] and territorial field sparrows [107], as well as the other vertebrates [108], suggesting that this neurotransmitter is mainly involved in empathy. Furthermore, it is known that serotonin enhances the release of vasopressin [109]: i.e., the two genes may exhibit epistasis.

Of course, the coding regions of the receptors are under strong selective pressure and should be identical within the groups. So they used a promoter region (SLC6A4 for HTTLPR), an intron (intron 2 VNTR for SLC6A4), as well as microsatellites (RS1 and RS3 for AVPR1a), which can be easily variable in a sampled population.

They also evaluated the dancers with the Tellegen Absorption Scale (TAS; a questionnaire that correlates positively with spirituality and altered states of consciousness) and the Reward Dependence factor in Cloninger's Tridimensional Personality Questionnaire (TPQ; a measure of need for social contact and openness to communication).

First, the dancers scored significantly higher in the two tests than the other two groups, suggesting that they tend to be more spiritual and social than the control groups. Second, they found significant differences in arginine vasopressin receptor polymorphism (*AVPR1a* haplotype frequencies of RS1 and RS3) between dancers and athletes according to a likelihood ratio test, especially when RS1 and RS3 were conditional on serotonin transporter (HTTLPR and VNTR) polymorphism. They obtained a similar result in comparison between dancers and nondancers/nonathletes. Third, TAS scores were also associated with the *AVPR1a* and *SLC6A4* haplotype frequencies.

Based on the results, they concluded that the two genes putatively associated with the dancing phenotype are rather involved in the emotional side of dance than the human sensory motor system. Their assertion is in fact very intriguing and provocative in terms of philosophical spiritualism: their conclusion implies that elite dancers control their bodies mainly by emotion, not by trained (i.e., "acquired") physical dexterity. In terms of molecular evolution, the implication is really deep.

On what ground they claimed so is the allele frequencies of the polymorphic data: i.e., they claimed that the emotion of the elite dancers might be associated with the allele frequencies, which are a seed of long-term evolution. In other words, the emotion of dancers potentially affects (or is affected by) allele frequencies of the haplotypes related to the two genes, and may drive their evolution in a way.

While the authors did not emphasize an evolutionary aspect of their study, the two dance genes may be able to be good markers to study human evolution, especially in terms of mental activities.

9 Evolutionary Perspectives on Dance

A professional dancer is required to express oneself through movement with flexibility, coordination, agility, gracefulness, a sense of rhythm, a feeling for music, and creativity [36, 103].

In this chapter, I showed several examples of dance-like behaviors performed by non-human animals, each of which has a different paradigm [110]. It is hard to directly compare between human and non-human dance-like behaviors due to the different paradigms. But still we can find the common key words between them: flexibility, coordination, agility, gracefulness, a sense of rhythm, a feeling for music, and probably creativity, especially for birds and primates. Even from invertebrates, we can easily draw key words shared with human: coordination and agility. In addition, nonverbal communication by dance-like movements is shared by many animals [111].

Despite of the diverged paradigms, it is obvious that vertebrates and invertebrates evolutionarily share basal neuromotor mechanisms [112]: while debate on evolution of centralized nervous systems is still underway [113], even fruit fly has a lot of characteristics of the neural system, which are stunningly similar to human [114].

When we regard dance as a phenotype, many things can be evolutionarily explained through genotypes, or genetic information. Fortunately, there are a lot of elaborate theories of molecular evolution and population genetics to analyze the genotypes [115]. And the genetic information is getting more available owing to the next generation sequencers (NGS) [116]. In this regard, for now, the most plausible way to carry out comparative analysis on "dance" is to obtain genetic information associated with the phenotypic behaviors, as Bachner-Melman et al. conducted.

Meanwhile, it is also possible to carry out a phenotype-driven approach. A few years ago, we have begun to apply a biomechanics framework to laboratory mouse, which is an excellent mammalian genetic tool [117]: i.e., we were trying to fuse the biomechanics with genetics. Recently, a group of Howard Hughes Medical Institute and Columbia University published an interesting paper somewhat similar to our framework [118], but with far more abundant funds. In near future, this kind of biomechanics-genetics framework will attract attentions of the research community, as one of promising interdisciplinary works [119].

An important issue here is that "phenotype" is still descriptive and often difficult to quantify [120, 121]. For example, how can we objectively quantify "coordination," "gracefulness," and "creativity?" While it should be complicated, the biomechanics framework may drop us a hint to handle the complexity.

At present, the most feasible way to analyze creative and spiritual characteristics on dance-related behaviors is to apply queries about TAS and TPQ to subjects. But we cannot use the same kinds of method to laboratory mice: they cannot answer to the queries. We need something else, and probably the biomechanics can be a good tool.

Regarding evolutionary perspectives on dance, comparative analysis between human and non-human analysis will provide critical insights, not only on neurosciences, but also on humanity: e.g., creativity. I believe that dance is an excellent framework to study human in terms of both mind and body.

Acknowledgements I would like to express my great appreciation to Professor Yoshihiko Nakamura of University of Tokyo. He introduced me the fascinating world of robotics, which was totally unknown to me. Professor Naruya Saitou of National Institute of Genetics, who is my PhD mentor, gave me insightful suggestions. This work is supported by Grants-in-Aid for Scientific Research—KAKENHI (26280110, 30391890, 30391890, and 17650122) from MEXT, Japan.

Glossary

Allele frequency Individuals in a population can have a variation of genes in each genomic location (locus). Each of the variation is called allele. An allele frequency is literally a proportion of the allele in the locus, among all the allele copies being considered.

Anatomically modern human It is equivalent to an individual member of the modern human in terms of paleoanthropology. Considering a possibility of the introgression between the modern human and *Homo* subspecies, this term is intentionally used: i.e., anatomically modern human may not be genetically modern human in terms of phylogeny.

Basal ganglia A region of brain involved in voluntary motor functions, procedural learning, eye movements, cognition, and emotion, which are all important for dance.

Central pattern generator (CPG) A (biological) neural network that generates rhythmic neural patterns as outputs without sensory feedback.

Coding region A genomic region that directly encodes amino acid sequences.

Darwinian selection In our context, Darwinian selection is virtually equivalent to the positive Darwinian selection (or directional selection): i.e., this is a mode of natural selection that favors an extreme phenotype, leading to a rapid shifting of the allele frequencies towards increment. This implies that the selected alleles have certain benefits to the individuals over the others. In the molecular evolution, the Darwinian positive selection is thought to be quite rare.

Effective population size The number of individuals in a population that contribute to the number of offspring in the next generation, which is usually smaller than the observed population size.

Epistasis A genetic concept that an effect of one gene depends on the presence of one or more modifier genes, especially in a non-additive way. The modifier genes are often called "genetic background."

Exogenous An adjective form of exogeny, which is phenomenon or an object originating externally: in biology, this often refers to DNA introduced to cells by transfection or vial infection.

Endogenous An adjective form of endogeny, which is an antonym of exogeny.

Fitness An individual's ability to propagate its genes in a population.

Genotype Discrete representation of a genetic trait of an individual, which is one of inherited determinants of a phenotype. The other determinants are typically epigenetic factors and non-inherited environmental factors.

Haplotype A collection of specific alleles in a cluster of tightly-linked genes on a chromosome. A set of single nucleotide polymorphisms (SNPs) is also called haplotype.

Intron Intermittent nucleotide sequences patched between exons in a coding sequence, which are removed by RNA splicing and do not contribute to encode proteins.

Microsatellite A tandem repeat of di-, tri-, or tetra-nucleotide units in genomic regions. The number of the repeats can easily vary during short-term evolution, leading to large polymorphism. So we can use the repeat number of a micorsatellite to identify a kinship.

Next generation sequencer Any high throughput sequencers capable of producing a huge amount of sequence data, by which we can exhaustively decode genetic information far quicker than by traditional methods.

Non-coding region An antonym of the coding region: i.e., DNAs that do not encode protein sequences. Some noncoding DNAs are transcribed to functional non-coding RNA sequences.

Phenotype Any characteristic traits (morphology, development, physiology, and behavior) of biological systems, which are determined by corresponding genotypes.

Polymorphic data Data representing polymorphism, which is a genetic or phenotypic variation in a population that shares a gene pool.

Phylogenetic tree A graphical representation of evolutionary processes by using a tree structure. Relationships of every extant and extinct organism can be represented with a single phylogenetic tree.

Population genetics Genetics that studies a process of allele frequency changes in a population rather than a individual-level phenomenon: i.e., natural selection, genetic drift, mutation, and gene flow.

Regulatory element A genetic component that regulates an expression of a gene or genes.

Retroviral sequence Endogeneous viral elements in the genome, which are highly homologous to known retroviruses. Surprisingly, they occupy 8 % of the human genome.

Selective sweep Reduction of variation in nucleotide sequences around a mutation that is subject to recent and strong positive selection.

Verbal dyspraxia A speech disorder, in which a person has difficulty in saying what she/he tries to express. Dyspraxia is motor function disability caused by an abnormality in the nervous system.

References

1. S. Chimpanzee, C. Analysis, Initial sequence of the chimpanzee genome and comparison with the human genome. Nature **437**, 69–87 (2005)
2. E.S. Savage-Rumbaugh, Verbal behavior at a procedural level in the chimpanzee. J. Exp. Anal. Behav. **41**, 223–250 (1984)
3. I. Roffman, E. Nevo, Can chimpanzee biology highlight human origin and evolution? Rambam Maimonides Med. J. **1**, e0009 (2010)
4. T. Romero, F.B. de Waal, Third-party postconflict affiliation of aggressors in chimpanzees. Am. J. Primatol. **73**, 397–404 (2011)
5. K. Harvati, The Neanderthal taxonomic position: models of intra- and inter-specific craniofacial variation. J. Hum. Evol. **44**, 107–132 (2003)
6. S.B. Carroll, Genetics and the making of Homo sapiens. Nature **422**, 849–857 (2003)
7. R.G. Klein, Darwin and the recent African origin of modern humans. Proc. Natl. Acad. Sci. **106**, 16007–16009 (2009)
8. S. Sankararaman, S. Mallick, M. Dannemann, K. Prufer, J. Kelso, S. Paabo, et al., The genomic landscape of Neanderthal ancestry in present-day humans. Nature **507**, 354–357 (2014)
9. R.E. Green, J. Krause, S.E. Ptak, A.W. Briggs, M.T. Ronan, J.F. Simons et al., Analysis of one million base pairs of Neanderthal DNA. Nature **444**, 330–336 (2006)
10. F. d'Errico, C.B. Stringer, Evolution, revolution or saltation scenario for the emergence of modern cultures? Philos. Trans. R. Soc. Lond. B Biol. Sci. **366**, 1060–1069 (2011)
11. A.L. Kaeppler, Dance in anthropological perspective. Ann. Revi. Anthropol. **7**, 31–49 (1978)
12. G.M. Morriss-Kay, The evolution of human artistic creativity. J. Anat. **216**, 158–176 (2010)
13. F.A. Issa, D.H. Edwards, Ritualized submission and the reduction of aggression in an invertebrate. Curr. Biol. **16**, 2217–2221 (2006)
14. J. Huxley, A discussion on ritualization of behaviour in animals and man. Philos. Trans. R. Soc. **251**, 249–271 (1966)
15. R. Huber, E.A. Kravitz, A quantitative analysis of agonistic behavior in juvenile American lobsters (Homarus americanus L.). Brain Behav. Evol. **46**, 72–83 (1995)
16. L. Gross, A new view of the waggle dance: making scents to recruit fellow foragers. PLoS Biol. **5**, e249 (2007)
17. R. Menzel, U. Greggers, A. Smith, S. Berger, R. Brandt, S. Brunke et al., Honey bees navigate according to a map-like spatial memory. Proc. Natl. Acad. Sci. USA **102**, 3040–3045 (2005)
18. K.V. Frisch, The dances of the honey bee. Bulletin of animal behaviour (translated from Die Tanze der Bienen (1946) Osterreiche Zoologie Zeitschrift 1:1–48) **5**, 1–32 (1947)
19. E. Crist, Can an insect speak? The case of the honeybee dance language. Soc. Stud. Sci. **34**, 7–43 (2004)
20. A.B. Barron, R. Maleszka, R.K. Vander Meer, G.E. Vander, Octopamine modulates honey bee dance behavior. Proc. Natl. Acad. Sci. USA **104**, 1703–1707 (2007)
21. J.R. Riley, U. Greggers, A.D. Smith, D.R. Reynolds, R. Menzel, The flight paths of honeybees recruited by the waggle dance. Nature **435**, 205–207 (2005)
22. S. Miyagawa, R.C. Berwick, K. Okanoya, The emergence of hierarchical structure in human language. Front Psychol **4**, 71 (2013)
23. A. Veeraraghavan, R. Chellappa, M. Srinivasan, Shape-and-behavior encoded tracking of bee dances. IEEE Trans. Pattern Anal. Mach. Intell. **30**, 463–476 (2008)
24. T. Landgraf, R. Rojas, H. Nguyen, F. Kriegel, K. Stettin, Analysis of the waggle dance motion of honeybees for the design of a biomimetic honeybee robot. PLoS ONE **6**, e21354 (2011)
25. J. Krause, A.F.T. Winfield, J.-L. Deneubourg, Interactive robots in experimental biology. Trends Ecol. Evol. **26**, 369–375 (2011)

26. R. Okada, H. Ikeno, T. Kimura, M. Ohashi, H. Aonuma, E. Ito, Error in the Honeybee Waggle dance improves foraging flexibility. Sci. Rep. **4** (2014)
27. B.E. Blasing, Segmentation of dance movement: effects of expertise, visual familiarity, motor experience and music. Front Psychol. **5**, 1500 (2014)
28. A. Goncearenco, I.N. Berezovsky, Prototypes of elementary functional loops unravel evolutionary connections between protein functions. Bioinformatics **26**, i497–i503 (2010)
29. T. Erclik, V. Hartenstein, R.R. McInnes, H.D. Lipshitz, Eye evolution at high resolution: the neuron as a unit of homology. Dev. Biol. **332**, 70–79 (2009)
30. D. Crews, The (bi)sexual brain. Science & society series on sex and science. EMBO Rep. **13**, 779–784 (2012)
31. D. Edwards, B. Chung, J. Bacque-Cazenave, D. Cattaert, Neuromechanical model of reflexes and locomotor rhythms in the crayfish leg. BMC Neurosci. **14**, 1–3 (2013)
32. S. Lavoué, M. Miya, M.E. Arnegard, P.B. McIntyre, V. Mamonekene, M. Nishida, Remarkable morphological stasis in an extant vertebrate despite tens of millions of years of divergence, in *Proceedings of the Royal Society of London B: Biological Sciences* (2010)
33. E.D. Jarvis, Learned birdsong and the neurobiology of human language. Ann. N. Y. Acad. Sci. **1016**, 749–777 (2004)
34. N.J. Emery, Cognitive ornithology: the evolution of avian intelligence. Philos. Trans. R. Soc. Lond. B Biol. Sci. **361**, 23–43 (2006)
35. N.J. Emery, N.S. Clayton, The mentality of crows: convergent evolution of intelligence in corvids and apes. Science **306**, 1903–1907 (2004)
36. N. Clayton, Q&A: bird behaviour, Darwin and dance. Interview by Patrick Goymer. Nature **462**, 288 (2009)
37. M.J. Fuxjager, K.M. Longpre, J.G. Chew, L. Fusani, B.A. Schlinger, Peripheral androgen receptors sustain the acrobatics and fine motor skill of elaborate male courtship. Endocrinology **154**, 3168–3177 (2013)
38. A. Hasegawa, K. Okanoya, T. Hasegawa, Y. Seki, Rhythmic synchronization tapping to an audio-visual metronome in budgerigars. Sci. Rep. **1**, 120 (2011)
39. R.A. Mulder, M.L. Hall, Animal behaviour: a song and dance about lyrebirds. Curr. Biol. **23**, R518–R519 (2013)
40. A.D. Patel, J.R. Iversen, M.R. Bregman, I. Schulz, Experimental evidence for synchronization to a musical beat in a nonhuman animal. Curr. Biol. **19**, 827–830 (2009)
41. A. Ravignani, D.L. Bowling, W.T. Fitch, Chorusing, synchrony, and the evolutionary functions of rhythm. Front Psychol. **5**, 1118 (2014)
42. A. Schachner, T.F. Brady, I.M. Pepperberg, M.D. Hauser, Spontaneous motor entrainment to music in multiple vocal mimicking species. Curr. Biol. **19**, 831–836 (2009)
43. M. Soma, C. Mori, The songbird as a percussionist: syntactic rules for non-vocal sound and song production in java sparrows. PLoS One **10**, e0124876 (2015)
44. H. Williams, Choreography of song, dance and beak movements in the zebra finch (Taeniopygia guttata). J. Exp. Biol. **204**, 3497–3506 (2001)
45. BBC, Planet earth: bird of paradise. Available: http://www.bbc.co.uk/programmes/p0036vcg (2009)
46. T. Shimizu, T.B. Patton, S.A. Husband, Avian visual behavior and the organization of the telencephalon. Brain Behav. Evol. **75**, 204–217 (2010)
47. M.I. Koukourakis, Radiation damage and radioprotectants: new concepts in the era of molecular medicine. Br. J. Radiol. **85**, 313–330 (2012)
48. J.W. Drake, Chaos and order in spontaneous mutation. Genetics **173**, 1–8 (2006)
49. M.W. Nachman, S.L. Crowell, Estimate of the mutation rate per nucleotide in humans. Genetics **156**, 297–304 (2000)
50. S. Kumar, S. Subramanian, Mutation rates in mammalian genomes. Proc. Natl. Acad. Sci. **99**, 803–808 (2002)
51. A. Poduri, G.D. Evrony, X. Cai, C.A. Walsh, Somatic mutation, genomic variation, and neurological disease. Science **341**, 1237758 (2013)

52. K.J. Peall, D.J. Smith, M.A. Kurian, M. Wardle, A.J. Waite, T. Hedderly et al., SGCE mutations cause psychiatric disorders: clinical and genetic characterization. Brain **136**, 294–303 (2013)
53. A.C. Antoniou, S. Casadei, T. Heikkinen, D. Barrowdale, K. Pylkas, J. Roberts et al., Breast-cancer risk in families with mutations in PALB2. N. Engl. J. Med. **371**, 497–506 (2014)
54. B. Kraszewska-Domanska, B. Pawluczuk, The gene of golden plumage colour linked with lower fertility in Mangurian Golden quail. Theor. Appl. Genet. **51**, 19–20 (1977)
55. F. Minvielle, S. Ito, M. Inoue-Murayama, M. Mizutani, N. Wakasugi, Genetic analyses of plumage color mutations on the Z chromosome of Japanese quail. J Hered **91**, 499–501 (2000)
56. S. Li, C. Wang, W. Yu, S. Zhao, Y. Gong, Identification of genes related to white and black plumage formation by RNA-Seq from white and black feather bulbs in ducks. PLoS One **7**, e36592 (2012)
57. S. Sato, T. Otake, C. Suzuki, J. Saburi, E. Kobayashi, Mapping of the recessive white locus and analysis of the tyrosinase gene in chickens. Poult. Sci. **86**, 2126–2133 (2007)
58. J. Barske, L. Fusani, M. Wikelski, N.Y. Feng, M. Santos, B.A. Schlinger, Energetics of the acrobatic courtship in male golden-collared manakins (Manacus vitellinus). Proc. Biol. Sci. **281**, 20132482 (2014)
59. J.J. Templeton, D.J. Mountjoy, S.R. Pryke, S.C. Griffith, In the eye of the beholder: visual mate choice lateralization in a polymorphic songbird. Biol. Lett. **8**, 924–927 (2012)
60. B.A. Schlinger, J. Barske, L. Day, L. Fusani, M.J. Fuxjager, Hormones and the neuromuscular control of courtship in the golden-collared manakin (Manacus vitellinus). Front Neuroendocrinol. **34**, 143–156 (2013)
61. M.L. Wilson, C. Boesch, B. Fruth, T. Furuichi, I.C. Gilby, C. Hashimoto et al., Lethal aggression in Pan is better explained by adaptive strategies than human impacts. Nature **513**, 414–417 (2014)
62. C. Boesch, C. Crockford, I. Herbinger, R. Wittig, Y. Moebius, E. Normand, Intergroup conflicts among chimpanzees in Tai National Park: lethal violence and the female perspective. Am. J. Primatol. **70**, 519–532 (2008)
63. R.W. Wrangham, M.L. Wilson, Collective violence: comparisons between youths and chimpanzees. Ann. N. Y. Acad. Sci. **1036**, 233–256 (2004)
64. R.W. Wrangham, L. Glowacki, Intergroup aggression in chimpanzees and war in nomadic hunter-gatherers: evaluating the chimpanzee model. Hum Nat **23**, 5–29 (2012)
65. M.N. Muller, S.M. Kahlenberg, M. Emery Thompson, R.W. Wrangham, Male coercion and the costs of promiscuous mating for female chimpanzees. Proc. Roy. Soc. Lond. B Biol. Sci. **274**, 1009–1014 (2007)
66. G.R. Pradhan, S.A. Pandit, C.P. Van schaik, Why do chimpanzee males attack the females of neighboring communities? Am. J. Phys. Anthropol. **155**, 430–435 (2014)
67. J. Goodall, Infant killing and cannibalism in free-living chimpanzees. Folia Primatol. **28**, 259–289 (1977)
68. H.M. Zawati, Impunity or immunity: wartime male rape and sexual torture as a crime against humanity. Torture **17**, 27–47 (2007)
69. G.G. Dimijian, Warfare, genocide, and ethnic conflict: a Darwinian approach. Proc. Bayl. Univ. Med. Cent. **23**, 292–300 (2010)
70. A. Papachristoforou, A. Rortais, J. Sueur, G. Arnold, Attack or retreat: contrasted defensive tactics used by Cyprian honeybee colonies under attack from hornets. Behav. Process. **86**, 236–241 (2011)
71. Y. Hattori, M. Tomonaga, T. Matsuzawa, Spontaneous synchronized tapping to an auditory rhythm in a chimpanzee. Sci Rep **3**, 1566 (2013)
72. F. Ramseyer, W. Tschacher, Nonverbal synchrony of head- and body-movement in psychotherapy: different signals have different associations with outcome. Front Psychol. **5**, 979 (2014)

73. N. Hugill, B. Fink, N. Neave, The role of human body movements in mate selection. Evol. Psychol. **8**, 66–89 (2010)
74. W.M. Brown, L. Cronk, K. Grochow, A. Jacobson, C.K. Liu, Z. Lalosević et al., Dance reveals symmetry especially in young men. Nature **438**, 1148–1150 (2005)
75. D. Garrigan, S.B. Kingan, M.M. Pilkington, J.A. Wilder, M.P. Cox, H. Soodyall et al., Inferring human population sizes, divergence times and rates of gene flow from mitochondrial, X and Y chromosome resequencing data. Genetics **177**, 2195–2207 (2007)
76. H. Li, R. Durbin, Inference of human population history from individual whole-genome sequences. Nature **475**, 493–496 (2011)
77. S.H. Ambrose, Late Pleistocene human population bottlenecks, volcanic winter, and differentiation of modern humans. J. Hum. Evol. **34**, 623–651 (1998)
78. A. Manica, W. Amos, F. Balloux, T. Hanihara, The effect of ancient population bottlenecks on human phenotypic variation. Nature **448**, 346–348 (2007)
79. S.L. Carto, A.J. Weaver, R. Hetherington, Y. Lam, E.C. Wiebe, Out of Africa and into an ice age: on the role of global climate change in the late Pleistocene migration of early modern humans out of Africa. J. Hum. Evol. **56**, 139–151 (2009)
80. J.J. Hublin, The earliest modern human colonization of Europe. Proc. Natl. Acad. Sci. USA **109**, 13471–13472 (2012)
81. F. d'Errico, M.A.F. Sánchez Goñi, Neandertal extinction and the millennial scale climatic variability of OIS 3. Q. Sci. Rev. **22**, 769–788 (2003)
82. W.E. Banks, F. d'Errico, A.T. Peterson, M. Kageyama, A. Sima, M.F. Sanchez-Goni, Neanderthal extinction by competitive exclusion. PLoS One **3**, e3972 (2008)
83. T. Wynn, F.L. Coolidge, *How to Think Like a Neandertal* (Oxford University Press, 2011)
84. T.M. Preuss, Human brain evolution: from gene discovery to phenotype discovery. Proc. Natl. Acad. Sci. USA **109**(Suppl 1), 10709–10716 (2012)
85. W. Enard, M. Przeworski, S.E. Fisher, C.S. Lai, V. Wiebe, T. Kitano et al., Molecular evolution of FOXP2, a gene involved in speech and language. Nature **418**, 869–872 (2002)
86. W. Enard, S. Gehre, K. Hammerschmidt, S.M. Holter, T. Blass, M. Somel et al., A humanized version of Foxp2 affects cortico-basal ganglia circuits in mice. Cell **137**, 961–971 (2009)
87. T. Maricic, V. Gunther, O. Georgiev, S. Gehre, M. Curlin, C. Schreiweis et al., A recent evolutionary change affects a regulatory element in the human FOXP2 gene. Mol. Biol. Evol. **30**, 844–852 (2013)
88. S. Mithen, *The Prehistory of the Mind: The Cognitive Origins of Art, Religion and Science* (Thames & Hudson, 1999)
89. B. Kurten, *Dance of the Tiger: A Novel of the Ice Age* (University of California Press, 1995)
90. V. Escott-Price, C. International Parkinson's Disease Genomics, M.A. Nalls, H.R. Morris, S. Lubbe, A. Brice, et al., Polygenic risk of Parkinson disease is correlated with disease age at onset. Ann. Neurol. **77**, 582–91 (2015)
91. J.F. Crow, On epistasis: why it is unimportant in polygenic directional selection. Philos. Trans. R. Soc. Lond. B Biol. Sci. **365**, 1241–1244 (2010)
92. M.S. Okun, Huntington's disease: what we learned from the original essay. Neurologist **9**, 175–179 (2003)
93. T. Steimer, The biology of fear- and anxiety-related behaviors. Dialogues Clin. Neurosci. **4**, 231–249 (2002)
94. M.C. Pfaltz, P. Grossman, T. Michael, J. Margraf, F.H. Wilhelm, Physical activity and respiratory behavior in daily life of patients with panic disorder and healthy controls. Int. J. Psychophysiol. **78**, 42–49 (2010)
95. M.J. Green, M.L. Phillips, Social threat perception and the evolution of paranoia. Neurosci. Biobehav. Rev. **28**, 333–342 (2004)
96. J.N. Crawley, *What's wrong with my mouse* (Wiley, New York, USA, 2000)
97. A. Ghosh, J. Rothwell, P. Haggard, Using voluntary motor commands to inhibit involuntary arm movements. Proc. Biol. Sci. **281**, 20141139 (2014)

98. A. Ghosh, P. Haggard, The spinal reflex cannot be perceptually separated from voluntary movements. J. Physiol. **592**, 141–152 (2014)
99. A. Walker, The strength of great apes and the speed of humans. Curr. Anthropol. **50**, 229–234 (2009)
100. G.H. Jacobs, Evolution of colour vision in mammals. Philos. Trans. R. Soc. Lond. B Biol. Sci. **364**, 2957–2967 (2009)
101. G. Jordan, J.D. Mollon, A study of women heterozygous for colour deficiencies. Vision. Res. **33**, 1495–1508 (1993)
102. T. Okayasu, T. Nishimura, A. Yamashita, O. Saito, F. Fukuda, S. Yanai et al., Human ultrasonic hearing is induced by a direct ultrasonic stimulation of the cochlea. Neurosci. Lett. **539**, 71–76 (2013)
103. R. Bachner-Melman, C. Dina, A.H. Zohar, N. Constantini, E. Lerer, S. Hoch et al., AVPR1a and SLC6A4 gene polymorphisms are associated with creative dance performance. PLoS Genet. **1**, e42 (2005)
104. T.R. Morrison, R.H. Melloni Jr, The role of serotonin, vasopressin, and serotonin/vasopressin interactions in aggressive behavior. Curr Top Behav Neurosci **17**, 189–228 (2014)
105. J. Borg, B. Andree, H. Soderstrom, L. Farde, The serotonin system and spiritual experiences. Am. J. Psychiatry **160**, 1965–1969 (2003)
106. C.F. Harding, S.A. Rowe, Vasotocin treatment inhibits courtship in male zebra finches; concomitant androgen treatment inhibits this effect. Horm. Behav. **44**, 413–418 (2003)
107. J.L. Goodson, Territorial aggression and dawn song are modulated by septal vasotocin and vasoactive intestinal polypeptide in male field sparrows (Spizella pusilla). Horm. Behav. **34**, 67–77 (1998)
108. E.A. Hammock, L.J. Young, Functional microsatellite polymorphism associated with divergent social structure in vole species. Mol. Biol. Evol. **21**, 1057–1063 (2004)
109. M. Galfi, M. Radacs, A. Juhasz, F. Laszlo, A. Molnar, F.A. Laszlo, Serotonin-induced enhancement of vasopressin and oxytocin secretion in rat neurohypophyseal tissue culture. Regul. Pept. **127**, 225–231 (2005)
110. C.C. Sherwood, F. Subiaul, T.W. Zawidzki, A natural history of the human mind: tracing evolutionary changes in brain and cognition. J. Anat. **212**, 426–454 (2008)
111. E. Genty, T. Breuer, C. Hobaiter, R.W. Byrne, Gestural communication of the gorilla (Gorilla gorilla): repertoire, intentionality and possible origins. Anim. Cogn. **12**, 527–546 (2009)
112. F. Lacquaniti, Y.P. Ivanenko, A. d'Avella, K.E. Zelik, M. Zago, Evolutionary and developmental modules. Front Comput. Neurosci. **7**, 61 (2013)
113. R.G. Northcutt, Evolution of centralized nervous systems: two schools of evolutionary thought. Proc. Natl. Acad. Sci. USA **109**(Suppl 1), 10626–10633 (2012)
114. F. Hirth, Drosophila melanogaster in the study of human neurodegeneration. CNS Neurol. Disord. Drug. Targets **9**, 504–523 (2010)
115. W.J. Ewens, F. James, Crow and the stochastic theory of population genetics. Genetics **190**, 287–290 (2012)
116. E.R. Mardis, Next-generation DNA sequencing methods. Annu. Rev. Genomics Hum. Genet. **9**, 387–402 (2008)
117. S. Oota, K. Mekada, Y. Fujita, J. Humphries, K. Fukami-Kobayashi, Y. Obata, et al., Four-dimensional quantitative analysis of the gait of mutant mice using coarse-grained motion capture. Presented at the Engineering in Medicine and Biology Society EMBC 2009. Annual International Conference of the IEEE (2009)
118. T. Akay, W.G. Tourtellotte, S. Arber, T.M. Jessell, Degradation of mouse locomotor pattern in the absence of proprioceptive sensory feedback. Proc. Natl. Acad. Sci. **111**, 16877–16882 (2014)

119. L.M. Havill, H.B. Coan, M.C. Mahaney, D.P. Nicolella, Characterization of complex, co-adapted skeletal biomechanics phenotypes: a needed paradigm shift in the genetics of bone structure and function. Curr. Osteoporos. Rep. **12**, 174–180 (2014)
120. C. Darabos, M.J. White, B.E. Graham, D.N. Leung, S.M. Williams, J.H. Moore, The multiscale backbone of the human phenotype network based on biological pathways. BioData Min. **7**, 1 (2014)
121. D. Papo, J.M. Buldu, S. Boccaletti, and E.T. Bullmore, Complex network theory and the brain. Philos. Trans. R Soc. Lond. B Biol. Sci. **369** (2014)

A Worked-Out Experience in Programming Humanoid Robots via the Kinetography Laban

Paolo Salaris, Naoko Abe and Jean-Paul Laumond

Abstract This chapter discusses the possibility of using Laban notation to program humanoid robots. Laban notation documents human movements by a sequence of symbols that express movements as defined in the physical space. We show, by reasoning around the simple action of "taking a ball", the flexibility of the notation that is able to describe an action with different level of details, depending on the final objective of the notation. These characteristics make Laban notation suitable as a high level language and as a motion segmentation tool for humanoid robot programming and control. The main problem in robotics is to express *actions* that are defined and operate in the physical space in terms of robot *motions* that originate in the robot motor control space. This is the fundamental robotics issue of inversion. We will first show how symbols used by Laban to describe human gestures can be translated in terms of actions for the robot by using a framework called Stack of Tasks. We will then report on an experience tending to implement on a simulated humanoid platform the notation score of a "Tutting Dance" executed by a dancer. Once the whole movement has been implemented on the robot, it has been again notated by using Laban notation. The comparison between both scores shows that robot's movements are slightly different from dancer's ones. We then discuss about plausible origins of these differences.

P. Salaris (✉) · N. Abe · J.-P. Laumond
LAAS-CNRS, 7 avenue du colonel Roche, 31000 Toulouse Cedex-4, France
e-mail: psalaris@laas.fr

N. Abe
e-mail: nabe@laas.fr

J.-P. Laumond
e-mail: jpl@laas.fr

J.-P. Laumond and N. Abe (eds.), *Dance Notations and Robot Motion*,
Springer Tracts in Advanced Robotics 111,
DOI 10.1007/978-3-319-25739-6_16

1 Introduction

1.1 Motions and Symbols

How to transform an action expressed in the physical space (i.e. "take the ball") in terms of a sequence of motions that originate in the motor control space (i.e. "bend the legs and then move the left hand forward")? The question constitutes the essence of robotics. It opens two main challenges: motion segmentation and motion generation. The segmentation of complex movements is a fundamental step in order to make easier programming the robot and execute a given action. However, the definition of these units of action or movement primitives remains an open problem. In that context, it is natural to consider dance notations as a promising route to decompose complex actions into sequences of elementary motions [1–3]. Indeed, the main purpose of dance notation is to store choreographic works and knowledge of dance techniques by translating movements taking place in physical space into specific ways as abstract symbols, letters, abbreviations, musical notations, stick figures, etc. In western culture, there are almost 90 dance notation systems, from the first appearance in 15th century to the present [4]. Today, among the most popular ones, we find the Kinetography Laban, the Benesh Movement Notation system and the Eshkol-Wachman Movement Notation system (see respectivelively Chalet-Haas', Mirzabekiantz' and Drewes' Chapters in this book).

This chapter reports on an experience in programming humanoid robots via the Kinetography Laban notation system.

The scope of Kinetography Laban is more general than only dance area. It aims at scoring all human motions independently of any behavior or any action. The system makes use of three types of movement symbols addressing respectively the direction of the movement, the part of the body doing the movement, the duration of the movement. It may be completed by a so-called "Effort" symbol describing the dynamic quality of the movement (see Loureiro de Souza' Chapter on Laban Movement Analyzis).

1.2 Action Versus Motion Segmentations. Physical Versus Control Spaces

All dance notation systems aim at describing the motions of the body as they are observed by human eyes. The purpose is to segment and to annotate the motions of body parts as expressed in the space (e.g. "the right hand is moving slowly forward"). The scoring operates in the physical space. A dancer who has been trained to dance notation is able to embody motion symbols: When reading a score, he/she "sees" the movement of the right hand in the physical space, and he/she "knows" what to do to move it slowly forward. Muscle activation is implicit. It is performed without explicit awareness of his/her muscle control space.

In robotics and in particular in humanoid robots, the segmentation of complex movements is a more complex issue. The robot has to obey a command given by an operator (e.g. “take the ball”). The command is expressed in the physical space as an action to be performed. The first difficulty is to express the command in terms of a sub-task sequence (e.g. “to take the ball, the robot has first to go to the ball and then to grasp the ball”). Tasks are then translate in terms of motion units (e.g. according to the context “take the ball” may mean “move the right hand forward”). Both the decomposition of an action into sub-tasks to be performed, and the translation of each sub-task in terms of motion units, constitute the first issue to be considered. The second issue is the translation of the motions expressed in the physical space into motions expressed in the control space. A robot does not a priori “know” what to do to move its hand forward. Differently from dance notations that take into account only the physical space around the body where actions are defined and operate, the issue of translating these actions, expressed in the physical space, into motions expressed in the motor control space, is fundamental in robotics.

Indeed, only once suitable control inputs are defined in the motor control space the robots can execute a given action in the physical space. However, the question of the segmentation is double. It deals with both action and motion decomposition:

- A given action may require a reasoning to decompose it into a sequence of sub-tasks to be performed. This is the task planning issue [5]. Figure 1 illustrates such a decomposition. To give Florent the ball, HRP2 robot has to locate the ball, to go to the ball, to take the ball, to locate Florent, to go to Florent, and finally to give Florent the ball. Elementary tasks as go to, take and give, require motion generation [6]. On the other hand, motion planning for robots manipulating movable objects among obstacles gives rise to decomposition issues [7]. The solution consists in structuring the configuration space of both the robot and the object into two elementary sub-spaces: the “grasping space” and the “placement space”. The topological structure of such sub-spaces directly reflects all possible segmentation of manipulation problems. The manipulation plan appears now as a sequence of motions lying in different sub-spaces that embed a natural decomposition of the problem, i.e. a natural segmentation that solves the manipulation planning. Task planning in such contexts is out of the scope of the current chapter.
- As motion segmentation is concerned, it is important to consider the clear distinction between physical space and motor control space. While dance notations operate movement segmentations in the physical space, robot programming requires to operate in the motor control space. The segmentation of a

Fig. 1 The action “give Florent the ball” is decomposed into a sequence of elementary tasks

movement in the physical space does not necessarily imply the same segmentation in the motor control space. For particular cyclic or repetitive actions, as e.g. elliptical and figure eight patterns of different sizes and orientations performed by using the whole arm, there is no evident segmentation in the motor control space [8] but rather continuous oscillatory patterns.

Coming back to Kinetography Laban, a Laban score represents by symbols *which* parts of the body, e.g. arms or legs, should move and *where*. The 27 direction symbols can be used to segment robot actions. These symbols can be translated in elementary tasks defined in the physical space. Each elementary task consists in moving a body part towards a desired direction specified by that symbol. The sequence of symbols then reproduces the whole action. Now the question for the roboticist is: how to translate motions defined in the physical space into motions defined in the motor control space? Indeed, Kinetography Laban is for humans and the brain can generate suitable signals to be sent to muscles in order to move arms, limbs and in general the whole body with the aim of reaching an establish final configuration. For humanoid robots the problem is much more challenging. First of all, these complex mechanical systems are usually actuated by motors instead of muscles. Each motor is in charge of moving a part of the robot with respect to another one. As a consequence, by controlling these motors it is possible to move the robot in order to make it walking, taking an object and in general executing established actions. In robotics, a method for controlling the motors in order to accomplish multiple tasks expressed in the physical space is the so called Stack of Tasks (SoT) [9].

1.3 Experience Overview

In this chapter, we show how the SoT can be used to translate the Laban score into control signals to be sent to the motors of a humanoid robot in order to move a body part towards a direction, specified by the corresponding symbol in the Laban score.

Section 2 introduces the SoT framework which basically provides a method for real-time control of redundant manipulators and hence useful also to control humanoid robots. We show how flexible is the notation to express human movements, and how this peculiarity is very useful in programing humanoid robots. While considering the example of the simple action of taking a ball on the floor, we show that Kinetography Laban allows annotating different levels of details.

In Sect. 3, we show how to translate Laban scores into tasks for the humanoid robot by using SoT. In particular, we will translate the Laban score of a “Tutting Dance”, which involves only upper body movements, in a hierarchical sequence of tasks in the SoT to be executed by a simulated humanoid robot (“Romeo” from Aldebaran Robotics). The final goal is to compare the Romeo’s movement with the dancer’s one. The main difference concerns the quality of the Romeo’s movement which is not same as the movement written in the original score.

For humanoid robots the problem of moving in a natural manner is very difficult and needs to answer the following question: which is the principle underlying the naturalness of humans' movements and hence what are the implicit rules that, after several years of human movement observations, are part of Kinetography Laban? In Sect. 4 we propose possible answers to this question on the basis of recent neuroscience and biological studies.

2 Robot Programming and Motion Notation: A Detailed Example

Let us consider the simple action of taking a ball. It might give rise to a complex motion involving the whole body and requiring the coordination of all body segments, if the ball is on the floor between the feet (see Fig. 2a). The legs have to naturally contribute to the action: "bending knees" becomes an integral part of the action "take the ball". In Fig. 2b the same action "take the ball" is programmed in

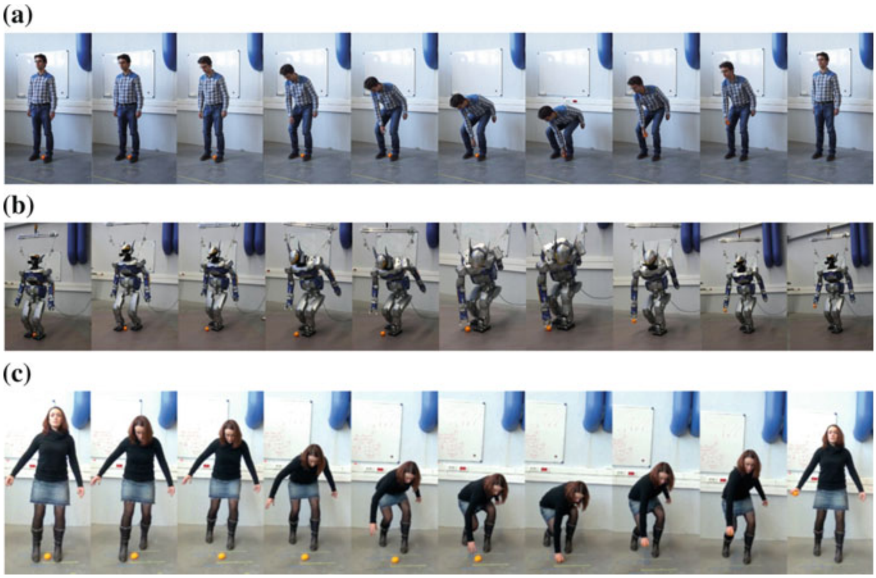


Fig. 2 To take the ball between its feet, contrarily to humans, the robot has to step away from the ball. This is due to the difference of morphologies between the human body and HRP-2 body. **a** Paolo takes the ball between its feet. **b** HRP-2 takes the ball between its feet. To do that the robot has to step away. In this experiment, "stepping away" is not a specific software module. It is not a symbol. It is an integral part of the embodied action of taking. The entire motion is the output of the sole and unique software module "take". **c** Tiphaine is executing a motion by reading the notation in Fig. 3c which describes the movements of HRP-2 in Fig. 2b when it takes the ball. Tiphaine does not a priori know that she is taking a ball

the humanoid robot HRP-2. Based on a simple comparison between Fig. 2a, b, it is straightforward to observe that humans and HRP-2 execute this action in different manners. Next subsections are dedicated to:

- describe how HRP-2 can be programmed in order to take the ball on the floor and explaining the differences between HRP-2 and humans engaged in the same action (Sect. 2.1), and
- how the flexibility of the notation in describing human movements can be exploited to notate with different level of details (see Sect. 2.2), going from directly translating the sentence “take the ball on the floor between feet” to precisely describing the movements of all body segments.

2.1 Robot Programming: The Stack of Tasks

One of the main feature of humanoid robots is their redundancy with respect to a task. This permits them to perform two or more tasks simultaneously, as e.g. taking an object with the right hand while putting the left one in touch with a fixed object to help legs ensure stability of the whole body. Of course, the robot can accomplish several tasks at the same time if and only if all of them are compatible each other: to check it, a simple idea is to put them in order.

The task-function approach [10], or operational-space approach [11], provides the mathematical framework to describe the hierarchy of tasks in terms of specific output functions, each one being a function from the configuration space to an arbitrary task space. At each time step of the integration process, a vector of the configuration space of the robot, tangent to the first task space, is selected. If the first task can be accomplished without using all the degrees of freedom of the robot, then a second task can be considered as soon it can be accomplished without interfering with the first one. By iterating this procedure, other tasks can be added until the whole set of degrees of freedom of the robot is completely exploited. The process can be iterated for other tasks so that a *stack of tasks* is obtained [9].

Let us now use the stack of tasks method with the humanoid robot HRP-2 in order to execute the action “take the ball” as showed by snapshots in Fig. 2a where a human subject is executing it. The required movement is made more complicated by the fact that the ball is between the feet. However, the action is very simple and consists only in reaching the ball with the end-effector of the right arm (or the left one) of the robot and taking the ball, while maintaining the static equilibrium. As a consequence, the tasks to be ordered in the stack of tasks are basically two. The first one consists in maintaining the static equilibrium that, on a horizontal ground, is verified when the projection of the center of mass of the robot is inside its support polygon—the footprint in case of single support phase. The second one, which has a lower priority in the stack of tasks, consists in zeroing the error between the current position of the end-effector of the right arm and the position of ball on the floor. However, this is not enough for a humanoid robot. Indeed, other kinematic constraints are necessary to enforce the joint limits and avoid self-collision.

A solution to this problem has been provided [12]. As shown in Fig. 2b the robot has to first step away from the ball and then grasp it. However, there is no dedicated module in charge of "stepping" and indeed, it is a direct consequence of "taking". The main reason why the robot step away from the ball can be found mainly in the kinematic constraint of avoiding self-collision. Moreover, by stepping away the robot reaches a position (the third snapshot in Fig. 2b) more comfortable for the robot, making easier the main task of maintaining the static equilibrium. The taking action is hence totally embedded in the body, allowing the legs to naturally contribute to the action. In this example, "taking" is an embodied action.

2.2 Motion Notation: The Kinetography Laban

The Kinetograhy Laban allows to write down not only dances but all kinds of human movements observable by human eyes. The utility of the Kinetography, outside of choreographic field, is mainly based on analyzing human actions. The Laban system describes a movement by four factors: space (by directions and levels), duration of the movement, beginning and end of the movement, and body parts. A complex action can be segmented according to these categorizations.

One of the basic element Laban notators use to write down a movement is the direction symbol. These symbols reflect a common approach to movement description in terms of spatial directions into which the part of the body move with the aim of reaching a given position [13]. The pathway taken is less important than the final destination. The Kinetography Laban is a movement notation because the symbols represent changes. As a consequence, an absence of movements is represented by an absence of symbols.

The directions in space emanate from a central point called *place*. It is represented by a rectangle. Directions and levels are computed from this point. There are 8 main directions and 3 shading levels to form 27 principal directions represented by modifications of the rectangle and by shading of each symbol. On their own the direction symbols state only information concerning the element of direction [13]. Only when they are placed in the appropriate column of the vertical staff it is possible to know which part of the body moves. In particular, for movements of the limbs direction and level are determined by the spatial relationship of the free-end (extremity of the limb) to the base (point of attachment). A line drawn between the extremity and the point of attachment indicates to which direction the limb has to move. The end can move with respect to the point of attachment, which is the point from which all the directions and levels specified by direction symbols radiate. For example, the whole arm is attached to the body by the shoulder. The shoulder is the point from which all directions and levels radiate. The whole arm can move with respect to the shoulder in order to place the hand, which is considered in this case as the free-end point of the arm.

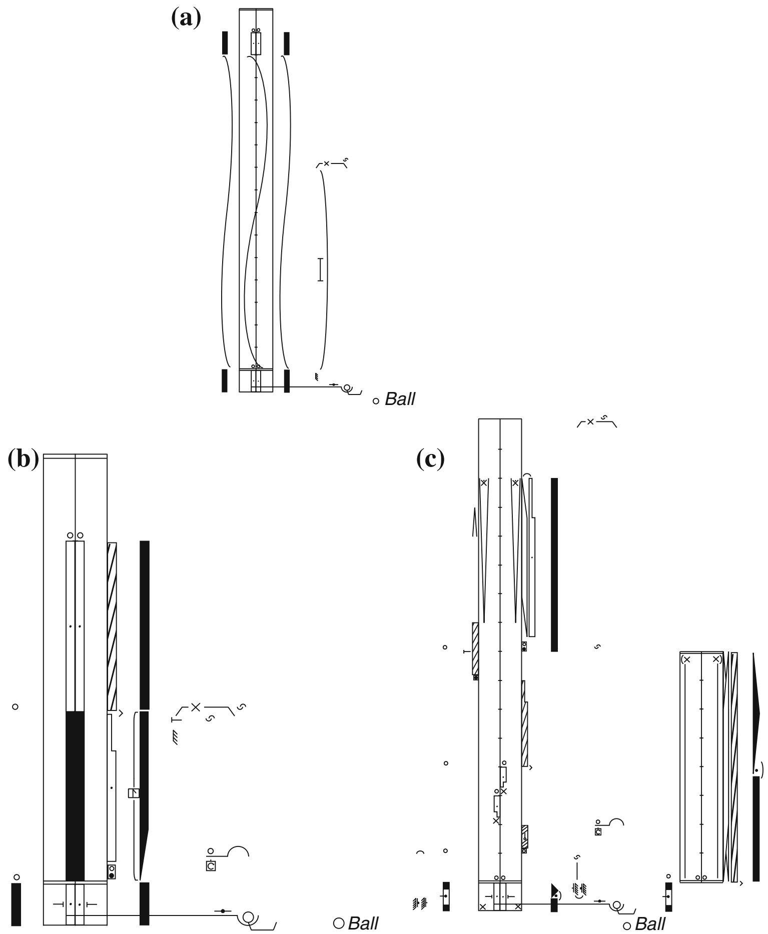


Fig. 3 Different Laban scores describing the motions motivated by the action of “Take the ball”. The figures may appear as obscure for readers not aware about Laban notation details. Their purpose is mainly illustrative to show that differences appear. Moreover, the presence of a symbol modeling the ball argues that the notation not only deals with the movement of the human body parts, but also with the movement of the ball. **a** Notation of the action “Take the ball” by using the Kinetography Laban. **b** The detailed description of the movement in Fig. 2a by using the Kinetography Laban. **c** The detailed description of the movement in Fig. 2b by using the Kinetography Laban

We notated the action of “taking the ball” in three manners. The notation of Fig. 3a is one of the simplest way among three to describe the action. This notation does not mention how to take the ball in detail, but it indicates only a starting position (standing), arms positions at the beginning and the end of the action (the arms are stretched out along the body), the position of the ball at the beginning (it means that the ball is on the floor between the feet.), and the right hand grasps the ball at a given moment. The only information included in the notation on the way of taking the ball is that the right hand follow a direct path to reach the ball. This is exactly the same information that is given to the robot via its programming system.

When we asked Paolo to take the ball without giving any constraints, he takes the ball without changing his feet positions. He just bends his knees and his hand grasps the ball. “Bending knees” is not explicitly expressed by the Laban score in Fig. 2a.

The score depicted in Fig. 3b is the notation of detailed Paolo’s movement. The notation describes his manner to take the ball with many details. It includes the way to reach the floor (e.g., rotation of the torso), the way to grasp (e.g. the choice of the right hand), the direction of the gaze, and a motion timing.

Figure 3 is a notation of the whole movement of HRP-2 robot. It includes exactly the same level of details as the score in Fig. 3b.

All the three scores differ. This is not a weakness of the notation system, but a strength. Indeed these three scores illustrate how the detail of an action can be described according to an intention of the notator. A same action may be noted with different levels of details according to the purpose of the notation, including what the notator wants to transmit to the performer and who is the reader of the score.

3 The Kinetography Laban as Robot Programming Tool

In this section, we show how the Laban score and in particular its 27 direction symbols can be translated in the framework of the Stack of Tasks. We will achieve this goal by a worked-out example where the simulated humanoid platform Romeo has to execute a kind of hip-hop dance, also known as “Tutting Dance”, that involves only the upper body, especially arms and hands, to create geometric shapes and movements.

3.1 The Tutting Dance: From Dancer Movements to the Laban Score

The Laban score of the Tutting Dance sequence[1] is shown in Fig. 4. The postures of the subject at each step of the Laban score are depicted in Fig. 5.

The score contains 9 columns. Each column is associated to a body part. The symmetrical stuff represents the symmetry of the body. The central columns on both sides of the main vertical axis represents the support of the body. Then the second columns are dedicated to the movement notation of the right and left legs. The third column, immediately outside the staff, are used for the torso and its parts. Tutting

[1]The original version of the Tutting dance can be found at the following link https://www.youtube.com/watch?v=082Akz8hGLY.

Dance mainly concerns the arms and the hands. This is why the second and third columns are free from any symbol, while the support columns contain only two small circles, which represent the “hold weight sign”, just after the double line which indicates the start of the movement. In this case, these symbols indicate that the actor has to maintain the standing posture and the weight on the feet.

The forth columns on the right and on the left of the body columns correspond to the right and left arm gestures, respectively. The fifth columns correspond to both forearms gestures, the sixth to both upper arms gestures, the seventh to the right and left hand gestures, the eighth to the back and palm of the right and left hands. Finally, the last columns on the right and on the left correspond to the edge of fingers. The duration of the sequence is decomposed into 16 intervals according to our movement segmentation. Moreover, based on our observations, we deduced that no movement is faster or slower than others. As a consequence, in the Laban score the direction symbols have the same length meaning that all the movements have the same duration.

A detailed description of each movement is here reported:

0. This is a starting point. An actor is standing. Both arms are stretched out along the body.
1. The right arm goes to the right-middle direction.
2. The left arm goes to the left-middle direction.
3. The right upper arm goes to the forward-middle. The right forearm goes to the left-middle direction.
4. The left upper arm goes to the forward-middle direction. The left forearm goes to the right-middle direction.
5. The left forearm goes to the forward-middle direction.
6. The right forearm goes to the forward-middle direction.
7. The left hand goes to down.
8. The right hand goes to up.
9. Both forearms go to up. During this movement, both hands maintain their configurations with respect to the forearms.
10. Both hands go to the right-middle direction.
11. The right upper arm goes to the left-forward-middle direction. The left upper arm goes to the right-forward-middle direction. Both upper arms are in contact. The left and right palms are also in contact.
12. The right and left hands while maintaining the contact, changed their direction to the left-middle.
13. Cancel the hold of the contact. The right upper arm goes to the intermediary direction between forward-middle and left-forward-middle. The left upper arm goes to the intermediary direction between forward-middle and right-forward-middle. As a consequence, the palms separate. The edge of the right finger touches the edge of the left hand.
14. The left hand goes to the right-middle direction. The left palm is in contact with the back of the right hand.

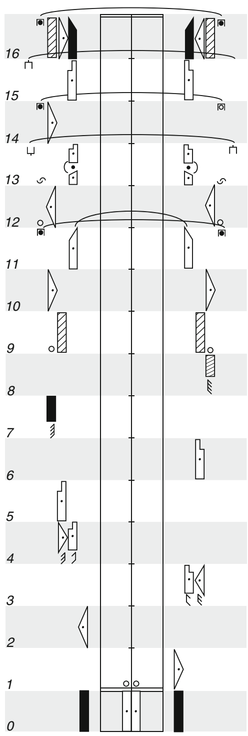


Fig. 4 Laban score for the Tutting dance

Fig. 5 Tutting dance

15. Both upper arms go to the forward-middle direction. The edge of both fingers are in contact.
16. The upper right arm goes to the right-forward-low direction. The right forearm goes to the left-middle direction. The upper left arm goes to the left-forward-middle direction. The left forearm goes to the right-middle direction. The hands go to up. Both palms are in contact.

3.2 From the Laban Score to Romeo Movements

Starting from the Laban score of the previous section (see Fig. 4), the Tutting Dance is now translated in the SoT so as to generate suitable control signals for the motors to execute the movements in the humanoid robot Romeo.

We have seen that one of the basic element Laban notators use to describe a movement is the direction symbol. The Tutting dance we have chosen leads itself very well to this type of approach due to its geometric shapes and movements.

The 27 principal direction symbols used to describe the Tutting Dance in Fig. 4 are the starting point to translate the Laban score in the SoT. In other words, depending on the current configuration and the body part which the symbol refers to, each principal direction symbol, and hence the main directions and levels, are translated as reference positions in space around the point of attachment. Each reference position is defined by an homogeneous transformation matrix that specifies both desired position and orientation of a reference frame attached at the free-end point of the movable part of Romeo. Based on the current position of the body part and the desired one specified by one of the principal direction symbol, a task function is defined as the error in terms of both rotation and translation between the current position in space of the reference frame attached to the free end and the desired one. The SoT is then used to determine suitable control signals for the motor of the robot such that this error becomes zero, while guaranteeing at the same time other tasks. These signals correspond to reference velocities or accelerations, the last ones providing a smoother movement for the robot. The first and most important task in humanoid robots is to maintain the static equilibrium. On a horizontal ground, the static equilibrium holds as soon as the projection of the center of mass of the robot is inside its support polygon—the footprint in case of single support phase. For the Tutting dance there is no displacement of the entire body (i.e. the so-called *weight* in Laban terminology and the so-called *root of the kinematic tree* in the robotics terminology) and indeed, in the support column of the Laban score, there are no symbols apart from the ones (small circles) that specify the maintenance of the starting feet positions. As a consequence, in the SoT, the first task, called “Weight on the feet” in Fig. 6, consists in guaranteeing the weight of the body rests on the feet.

We have seen that the absence of movements is represented by the absence of symbols. To include this rule in the humanoid robot Romeo, the last task in the

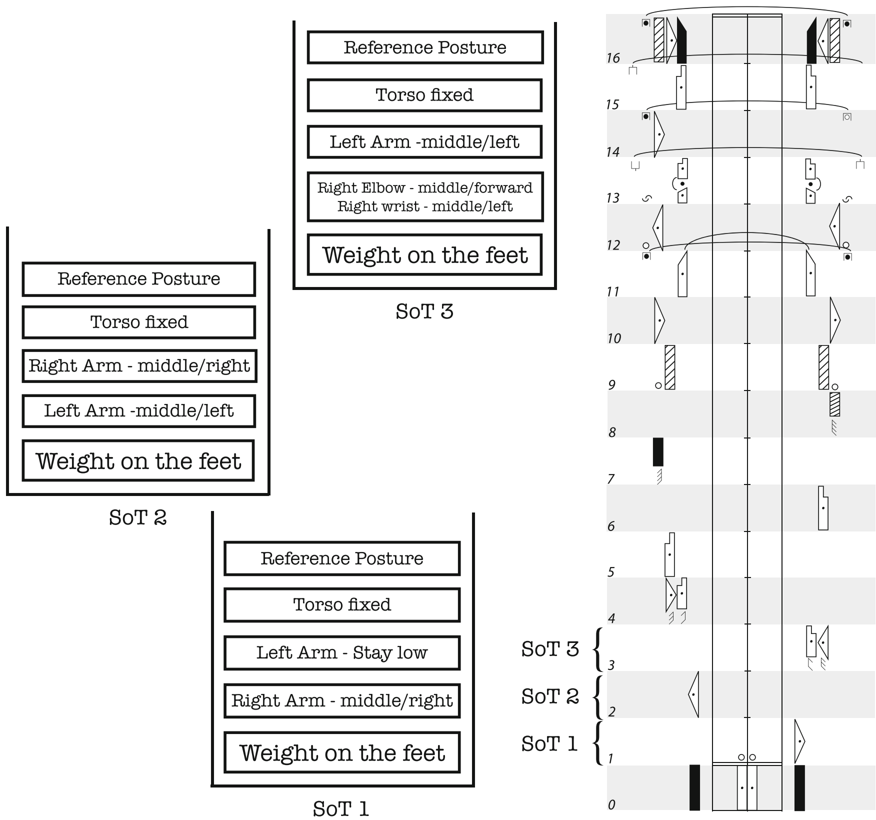


Fig. 6 Stack of tasks for the first tree steps of the Laban score

stack, called “Reference posture” in Fig. 6, consists in limiting the “distance” from a reference configuration that can be considered as the natural standing position for humans. Just before this task, so at higher priority, a task, called “Torso fixed” in Fig. 6 is added in order to account for the lack of symbol in the body columns. Between the first task “Weight on the feet” and the task “Centre of Mass”, other tasks representing the gesture symbols that follow one another in the Laban score will be added as represented in Fig. 6.

Another rule of the Kinetography Laban is that, after each symbol, until a new symbol does not involve the same part of the body (e.g. the whole arm), or a sub-part of it (e.g. the forearm), the previous symbol holds. For this reason, in the stack of task some tasks change priority before disappearing from the stack: for example, referring to Fig. 6, task “Right Arm—middle/right” has a lower priority in SoT 2 with respect to SoT 3.

The final result of the implementation of the Tutting dance in Romeo is shown in Fig. 7. Next section is dedicated to show the main differences between Romeo and human movements.

4 Discussion

In this section, we will first discuss on the obtained results, by comparing the movement of Romeo with the movement of Naoko, also by means of the Kinetography Laban. The second subsection is dedicated to describe some rules of the Kinetography Laban about how movements should be executed, and plausible origins of these rules.

4.1 Comparison Between Human and Romeo Movements

In the previous section, the Laban score of a simple example of Tutting Dance has been translated, by using the SoT, in suitable control signals for the motor of the simulated humanoid robot Romeo. In Fig. 7, the snapshots corresponding to the Laban score in Fig. 5 is reported. Apart from the standing posture of Romeo which is slightly different from humans (the legs are slightly bent, as most humanoid robots), there are not significant differences w.r.t. the snapshots in Fig. 5. However, in snapshot 3 of Fig. 7, the left arm of Romeo is bent while the left arm of Naoko is straight (see snapshot 3 of Fig. 5). Moreover, mainly in the asymmetric postures,

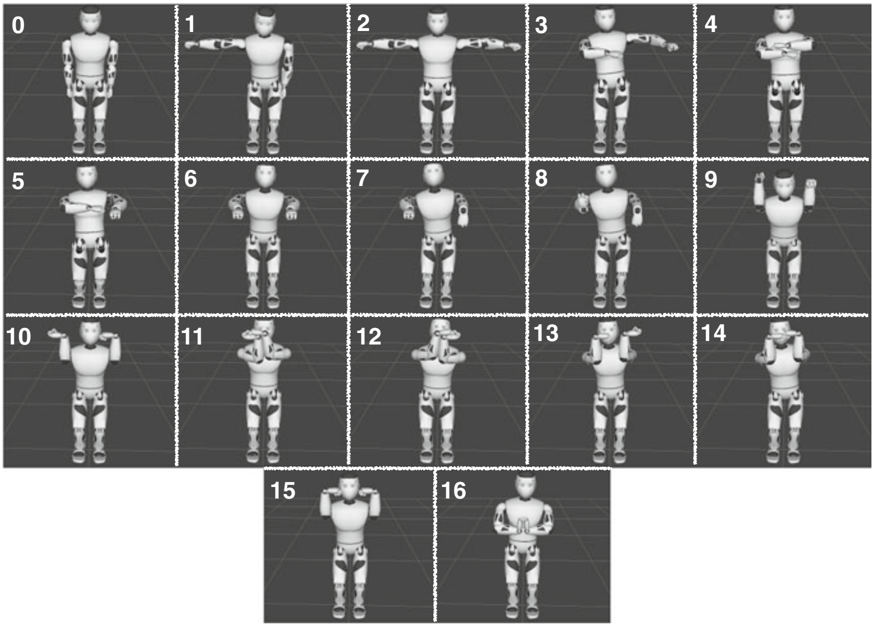


Fig. 7 Tutting dance realized by Romeo. The numbering of each snapshot corresponds to the numbering in the Laban score of Fig. 4

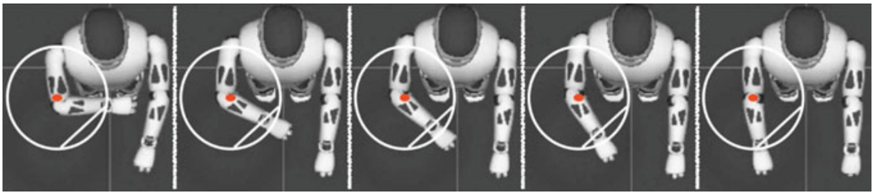


Fig. 8 Referring to Fig. 4, this is a sequence of snapshots for the movement at step 6 in the Laban score of Fig. 4. The free end of the right arm moves along a straight line between the starting position to the final one. Hence, the elbow (*red point*) does not remain at a fixed position in space. This makes the difference with the original motion performed by Naoko

the torso, the pelvis and shoulders of Romeo are not exactly at the same configurations of Naoko at the end of each step.

All these different configurations at the end of each step are a direct consequence of the movement executed by Romeo while progressing in the Laban score. Indeed, by using the SoT to generate the movements, each motor of Romeo is controlled such that the free end of a body segment moves along a straight line passing from both the initial position and the final one, as represented in Fig. 8. The movement corresponds to the step 6 in the Laban score. The free end of the right forearm should move following a circular arc centered at the elbow, i.e. performing a peripheral movement, and the rest of the body should remain fixed. Based on the snapshots in Fig. 8 this is not the case, as it is also pointed out in Fig. 9.

This characteristic gives hence rise to undesirable movements of the body. In Fig. 9a, the movement corresponding to the first step in the laban score and to SoT 1 in Fig. 6 is shown. All these differences can be also appreciated by notating the Romeo movement with the Kinetography Laban. In Fig. 10 a comparison between the Laban scores of Romeo's and dancer's movement during the first 6 steps is reported. Notice that now, in the Laban score of Romeo (see Fig. 10b), the columns for the body gestures is not free of symbols and signs. Moreover, on the side of each direction symbol, a new sign is notated, representing a description of the path that the free end of the arm is now executing—basically, a straight line. In Fig. 9b, snapshots corresponding to the movement from step 2 to step 3 of the Laban score is also reported. In this case, the movement of the whole body is much more noticeable.

Finally, even if the Laban score in Fig. 10a is a so called *legato movement*, that is with no separation between direction symbols and hence with no interruption in the continuity of the movement, for Romeo, sometimes there is a sort of overlapping between consecutive direction symbols and hence movements that indeed should be consecutive. This means that still before the previous movement is finished, the next one is already started (see Fig. 10b between steps 3 and 4).

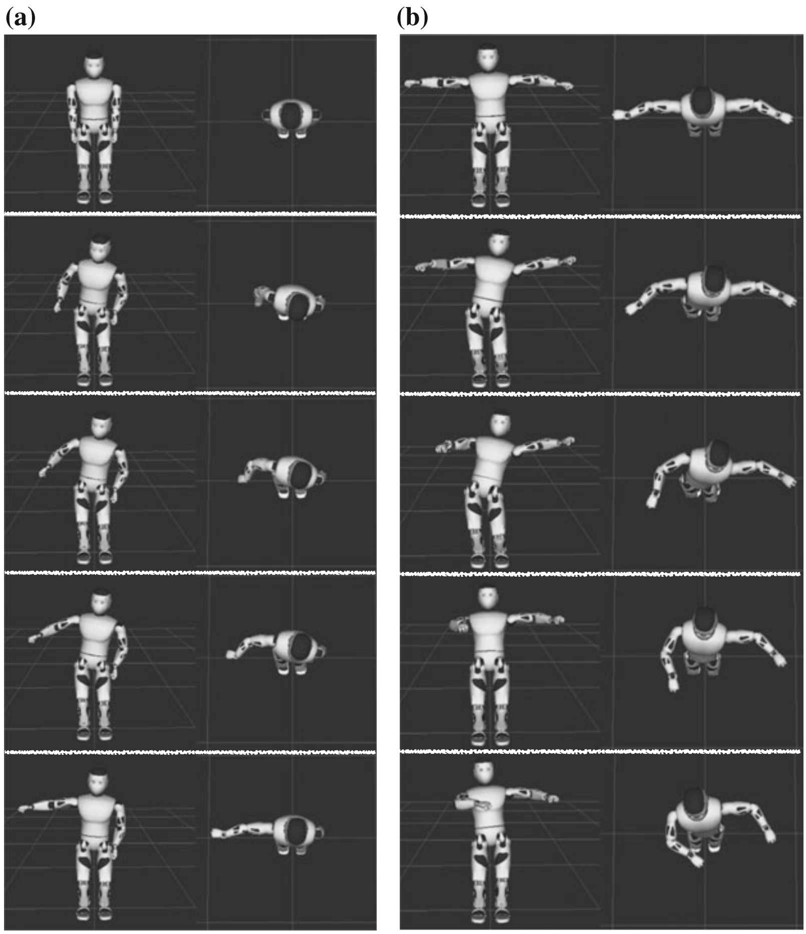


Fig. 9 Two movements executed by Romeo. All movements describe straight line paths for the free end of the arms. This gives rise to undesirable movements of the body. **a** The movement of Romeo at step 1 in Fig. 4. **b** The movement of Romeo at step 3 in Fig. 4

4.2 On the Naturalness of Movements

As already said in previous section, on their own the direction symbols state only information concerning the element of direction. Only when they are placed in the appropriate column of the vertical staff it is possible to know which part of the body has moved. Moreover, depending on the actual configuration of the body part, due to implicit rules of the Kinetography Laban created and based on the naturalness of human movements, information on the path of the free end of that part of the body can be also achieved.

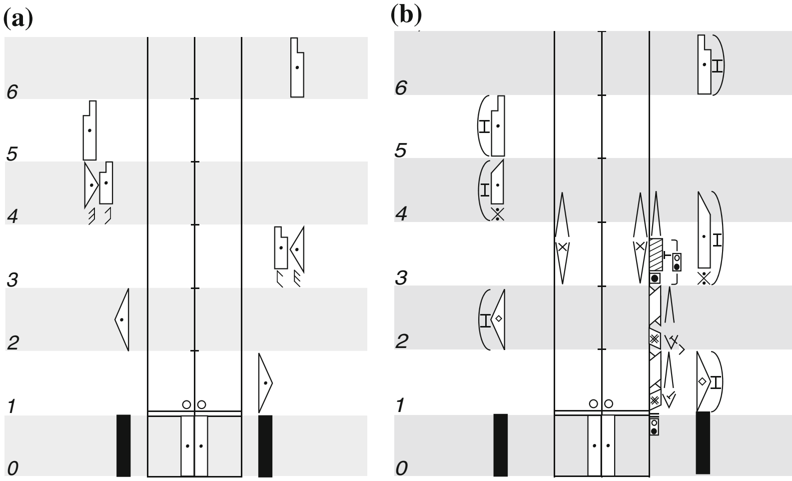


Fig. 10 A comparison between the Laban score of the Tutting Dance executed by a dancer and Romeo. **a** Laban score for a dancer. **b** Laban score for Romeo

In using the 27 principal directions around the body, symbols that correspond to adjacent points in space are considered to be at a *first-degree distance* from one another (see Fig. 11). For example, if the arm moves from forward middle to the adjacent right front diagonal point, this is a first-degree distance. In this case, the free end of the arm, i.e. the hand, describes an arc of circle whose center is the shoulder moves on the surface of the sphere. This is the so called *peripheral movement* in Laban system [14]. All movements between first-degree distance points will produce this path. In case of the arm moves from forward middle to side (right or left) middle, we have a *second-degree distance* (see Fig. 11) and the hand describes a quarter of circle whose center is the shoulder, it is also called *peripheral movement*. All movements between second-degree points have to be performed without any special flexion of the arm unless otherwise specified with the addition of a particular sign, as e.g. the straight path sign.

If two points are at *third-degree distance*, the free end of the limbs moves along a trajectory closed to the body, and not peripheral path. In this movement, the arm is slightly bent, takes a path between periphery and center ("*in place*"). This is called *intermediate situation or transversal movement* [14].

Finally, diametrically opposite points are considered to be at a fourth-degree distance. For example if the arm moves from forward middle to the extreme opposite direction backward middle. The arm comes back "*in place*" then extends again to outside. These type of movements are called *central movement* in Laban system [14].

The control laws used to move the arms of Romeo do not contain all this variety of movements. Independently from the distance of points in space, Romeo always performs a straight line path, with noticeable loss in naturalness. Indeed, to perform that movement, other undesirable and unnatural body movements are necessary. Of

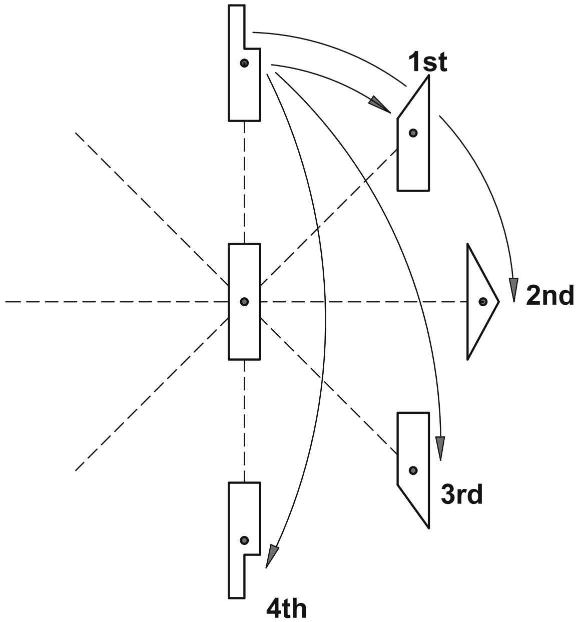


Fig. 11 Degree of distance between points

course, to solve this problem, an ad-hoc control law that moves the hand along peripheral, central, or transversal movement might be determined.

On the other hand, the implicit rule of the notation comes from the several years of observations and it is based on the naturalness of human movements, also induced by the mechanical structure of the body. As often conjectured in robotics, an optimality principle might underline human movements. It would be hence interesting to understand what is this principle, to express it in a suitable mathematical manner and than to use it to determine the control laws for the humanoid robot Romeo, by using tools as the optimal control theory.

The main problem in translating the Laban score in humanoid robots is hence to obtain the continuous movement that resemble the human one. To this end, several biological studies show principles that explain human movements [15–18]. In particular, considering only arm movements in an horizontal plane, the recorded trajectories are well explained in terms of minimum jerk [15]. On the other hand, when arms move on a vertical plane (in this case the force of gravity plays a fundamental role, making the movement asymmetric), the recorded trajectories are well explained in terms of the minimum sum of jerk and energy [18].

5 Conclusion

What are the lessons learned from this attempt to make use of Kinetography Laban in humanoid robot programming?

Three study years are necessary to be graduated in Laban notation. Even for well trained notators, scoring simple movements as the ones described in this chapter is time consuming. We have seen that a same action may be notated with different levels of details. Indeed Laban notation addresses the movement more than the action itself. Two scores may symbolize the action of "taking a ball" according to the importance one gives to the manner to take a ball. Both scores are complementary; they are not opposite. For a roboticist, making use of Laban notation to program a humanoid robot requires either to have been trained for several months, or to work with a notator. At the very end, the notator describes the movements in the physical space and the roboticist translates the score symbols in terms of Jacobian inversion. More the score is detailed, more the inversion task is tedious. This makes the question of automated translation a critical issue. In spite of few tentatives referenced in the introduction of this paper, the question remains largely open, as soon as the objective is to account for all dimensions and richness of the Kinetography Laban notation.

Human body and humanoid robot body differ. Retargeting a human movement on a digital artifact is a well known issue in computer graphics [19]. This is the same in humanoid robotics. We have seen that HRP-2 robot cannot take the ball as Paolo did. This is due to the fact that the body of HRP-2 prevents the robot to put its hand between its feet without stepping away. To take the ball on the floor, HRP-2 robot may benefit from the score in Fig. 2b but not from the score in Fig. 2c. Human motion notations are all based on the kinematic structure of the human body. Adapting the notation to another structure is certainly possible, but it is a challenge by itself.

Finally we have seen that the question of the naturalness of a movement is another critical issue. Laban notators benefit from a lot of implicit knowledge that ground the Kinetography Laban. The issue has been clearly revealed in Sect. 4.2. The rules to move the hand in a given direction have been defined and described on the basis of a long experience in observing human movements. Movement notators target a movement description. They are not a priori interested by causality principles, i.e. by the origin of the movement. The origin of the movement takes place in the muscle control space. However, it is not necessary to tell a dancer what muscles he/she has to activate to perform a desired movement. Muscle activation is an unconscious process. With the fundamental problem of inverting actions expressed in the physical space into motor controls, roboticists have to face the causality principle. This is why, like neurophysiologists, roboticists try to exhibit general movement laws to explore plausible causality principles.

To summarize the experience gained with the worked-out examples described in this paper, we can say that dance notation and robot programming pursue two different goals in two different spaces. The goal of the dance notator is to describe the qualities of the movement as wished by the choreographer while the roboticist is a priori concerned by the action to be performed, better than by the motion that fulfills the action. However we have seen that there is an interest for roboticists to consider dance notation as a way to better explore the relationship between action and motion. Symbolical and computational foundations of both motion and action concepts, as complementary developed by dance notation practitioners and roboticists respectively, deserve to be deeper explored.

Acknowledgements This work is supported by ERC-ADG project 340050 Actanthrope. Authors thank Noëlle Simonet, professor of the Kinetography Laban at CNMDP (Conservatoire National de Musique et de Danse de Paris), for reviewing the Laban scores, and Tiphaine Jahier, dancer and Laban notator, for her participation to read notations and perform actions.

References

1. B. Choi, Y. Chen, *Humanoid Motion Description Language* (Lund University Cognitive Studies, 2002)
2. K. Kahol, P. Tripathi, S. Panchanathan, Automated gesture segmentation from dance sequences, in *2004 Proceedings of Sixth IEEE International Conference on Automatic Face and Gesture Recognition* (IEEE, 2004), pp. 883–888
3. K. Kahol, P. Tripathi, S. Panchanathan, T. Rikakis, Gesture segmentation in complex motion sequences, in *Proceedings IEEE International Conference on Image Processing* (2003), pp. 105–108
4. A. Hutchinson Guest, *Choreo-Graphics, A Comparison of Dance Notation Systems From the Fifteenth Century to the Present* (Gordon and Breach Science Publishers S.A., New York, 1989)
5. M. Ghallab, D. Nau, P. Traverso, *Automated Planning: Theory and Practice* (Morgan Kaufmann Publishers, Elsevier, 2004)
6. E. Yoshida, J.P. Laumond, C. Esteves, O. Kanoun, A. Mallet, K. Sakaguchi, T. Yokoi, Motion autonomy for humanoids, experiments on hrp-2 no. 14. Comp. Animation Virtual Worlds **20**, 5–6 (2009)
7. T. Siméon, J.-P. Laumond, J. Cortés, A. Sahbani, Manipulation planning with probabilistic roadmaps. Int. J. Robot. Res. **23**(7–8), 729–746 (2004)
8. D. Sternad, S. Schaal, Segmentation of endpoint trajectories does not imply segmented control. Exp. Brain Res. **124**(1), 118–136 (1999)
9. N. Mansard, F. Chaumette, Task sequencing for high-level sensor-based control. IEEE Trans. Rob. **23**(1), 60–72 (2007)
10. C. Samson, B. Espiau, M. Le Borgne, *Robot Control: The Task Function Approach* (Oxford University Press, Oxford, 1991)
11. O. Khatib, A unified approach for motion and force control of robot manipulators: the operational space formulation. Robot. Autom. IEEE J. **3**(1), 43–53 (1987)
12. O. Kanoun, J.-P. Laumond, E. Yoshida, Planning foot placements for a humanoid robot: a problem of inverse kinematics. Int. J. Robot. Res. **30**(4), 476–485 (2011)
13. A.H. Guest, *Labanotation: The System of Analyzing and Recording Movement* (Psychology Press, 2005)
14. A. Knust. *Dictionnaire usuel de Cinétographie Laban (Labanotation)*. Ressouvenances (2011)
15. T. Flash, N. Hogan, The coordination of arm movements—an experimentally confirmed mathematical-model. J. Neurosci. **5**(7), 1688–1703 (1985)
16. P.J. Stapley, T. Pozzo, G. Cheron, A. Grishin, Does the coordination between posture and movement during human whole-body reaching ensure center of mass stabilization? Exp. Brain Res. **129**(1), 134–146 (1999)
17. A. Sciutti, L. Demougeot, B. Berret, S. Toma, G. Sandini, C. Papaxanthis, T. Pozzo, Visual gravity influences arm movement planning. J. Neurophysiol. **107**(12), 3433–3445 (2012)

18. J. Gaveau, B. Berret, L. Demougeot, L. Fadiga, T. Pozzo, C. Papaxanthis, Energy-related optimal control accounts for gravitational load: comparing shoulder, elbow, and wrist rotations. J. Neurophysiol. **111**(1), 4–16 (2014)
19. R. Parent, *Computer Animation: Algorithms and Techniques* (Morgan Kaufmann Publishers, Elsevier, 2001)

Using Dynamics to Recognize Human Motion

Gentiane Venture, Takumi Yabuki, Yuta Kinase, Alain Berthoz and Naoko Abe

Abstract We explore the importance of the dynamics of motion, and how it can be used first to develop and to personalize intelligent systems that can understand human motions, then to analyze motions. We propose a framework that uses not only the kinematics information of movements but also the dynamics and allows to classify, analyze and recognize motions, emotions in a non-verbal context. We use the direct measure of the dynamics when available. If not we propose to compute the dynamics from the kinematics, and use it to understand human motions. Finally, we discuss some developments and concrete applications in the field of motion analysis and give some experimental results using gait and simple choreography.

1 Introduction

Human motion science and motion analysis have a long history [1]. The Renaissance has been a fruitful period in understanding the underlying mechanics of the human body, and it is in the 17th century that Isaac Newton, and later Lagrange, have really enabled to write the equations of motion as we know them. The 19th century and the 20th century have then enable to measure what these

G. Venture (✉) · T. Yabuki · Y. Kinase
Tokyo University of Agriculture and Technology, 2-24-16 Nakacho, Koganei, Tokyo, Japan
e-mail: venture@cc.tuat.ac.jp

A. Berthoz
Laboratoire de Physiologie et de Perception de l'Action, College de France,
11 Place Marcelin Berthelot, 75005 Paris, France
e-mail: alain.berthoz@college-de-france.fr

N. Abe
Gepetto TEAM, LAAS-CNRS, 7 Avenue du Colonel Roche, 31400 Toulouse, France
e-mail: nabe@laas.fr

J.-P. Laumond and N. Abe (eds.), *Dance Notations and Robot Motion*,
Springer Tracts in Advanced Robotics 111,
DOI 10.1007/978-3-319-25739-6_17

equations were expressing, with the development of photography and motion picture and the famous work of Muybridges and Marey to cite only the most famous.

The last decade of the 20th century and the 21st century have been marked by the developments of motion capture technology, with faster capture rates, higher measuring precision and resolution. Not only can the kinematics of a motion be measured but its dynamics too with the use of force sensors and force-plates. These systems have paved the way for numerous applications and developments and are now commonly used in animation, sport science, motor rehabilitation and robotics [2–4]. On the other hand dance and performance professionals have developed powerful systems to annotate, record and reproduce extremely complex choreographies and motion sequences [5]. From the kinematics information, systematic annotation can be done [6], in addition from a set of notations, motion kinematics can be generated and therefore used for animation [7]. These offer an extremely powerful mean to generate and record motion of any entities, human or not and the annotation are crucial to understand these motions. Labanotation (also known as Kinetography Laban) offers a compact and powerful notations system for motions. Each motion can be described with a number of defined elements just like a score in music. Labanotation treats of the dynamics which is expressed by the body of a performer, and the way the performer interacts with the environment which is a difficult concept. However Labanotation does not quantify or measure these factors. Yet the dynamics is an important aspect of motion since entities are bound to external forces: gravity, and then interact with their environment to create motion, to generate more or less force, express something, convey energy. The sole kinematics cannot describe these interactions and thus incompletely describes motions. Our research offers a mean to quantify and bridge the gap between classical dynamics analysis in biomechanics and in robotics and in annotations in Labanotation.

We present a general framework to analyze and classify motions and in particular dynamics in order to highlight some specific value: motion, emotion, or mood that can be used to systematically define dynamics as defined in Labanotation [8]. Our framework is based on the definition of a feature vector of the data and of a decomposition in principal components (PC). The decomposition in the PC space is used as the base of our classification and recognition algorithm.

The chapter is organized as follows: in Sect. 2, we first describe the terminology used to describe a motion and its dynamics and how it relates to or differs from Labanotation developed to analyze and record motion in dance and more generally performances. Section 3 presents the feature vector analysis using principal component analysis (PCA). Section 4 presents concrete examples using torso data and contact forces. Section 5 provides some extension of the results and proposes to use dynamics information for emotion and mood classification. Section 6 concludes this chapter by offering some perspectives for future applications and developments.

2 Equation of Motion and Terminology

The equation of motion (1) in its most general form [9, 10] provides a simple mean to understand the time variant relationship between what is called in Biomechanics [1] and Robotics [11] a motion: displacement and orientation of a reference, joint angles of the articulated system; the forces and moment of forces applied on the environment and the forces generated by the muscles and transmitted through the joint torques. It also makes use of time invariant such as the body segment lengths and the dynamic properties of these segments, such as masses and inertia.

$$\begin{bmatrix} \mathbf{0} \\ \boldsymbol{\Gamma} \end{bmatrix} + \begin{bmatrix} \boldsymbol{J}_b^T \\ \boldsymbol{J}^T \end{bmatrix} \boldsymbol{F} = \begin{bmatrix} \boldsymbol{M_b} & \boldsymbol{M_c} \\ \mathbf{0} & \boldsymbol{M_c} \end{bmatrix} \begin{bmatrix} \ddot{\boldsymbol{q}}_b \\ \ddot{\boldsymbol{\theta}} \end{bmatrix} + \begin{bmatrix} \boldsymbol{C_b} & \boldsymbol{C_c} \\ \mathbf{0} & \boldsymbol{C_c} \end{bmatrix} \begin{bmatrix} \dot{\boldsymbol{q}}_b \\ \dot{\boldsymbol{\theta}} \end{bmatrix} + \begin{bmatrix} \boldsymbol{G_b} \\ \boldsymbol{G_c} \end{bmatrix} \tag{1}$$

where $\boldsymbol{M_c}$, $\boldsymbol{C_c}$, $\boldsymbol{G_c}$, $\boldsymbol{M_b}$, $\boldsymbol{C_b}$, and $\boldsymbol{G_b}$, are the inertia, Coriolis and gravity matrices calculated at the articulated chain (subscript c) and at the floating base level (subscript b), respectively. $\dot{\boldsymbol{\theta}}$, $\ddot{\boldsymbol{\theta}}$, and $\boldsymbol{\Gamma}$ are the joint velocity, acceleration and torque vectors, respectively. $\boldsymbol{J}^T$ is the Jacobian transpose matrix mapping the external forces to the joint space and $\boldsymbol{J}_b^T$ is the Jacobian transpose of the matrix mapping the external forces to the floating base frame. Finally, $\dot{\boldsymbol{q}}_b$, and $\ddot{\boldsymbol{q}}_b$ are the base Cartesian velocity and acceleration vectors, respectively.

Depending on the available measures and the region of interest for the human motion analysis different approaches can be considered. Some studies [12, 13] consider only the movement of the extremities and or the torso since they are of low dimensions, what is often called the *task space* in robotics [11]. Labanotation system describes the displacement of weigh (*weight transfer*) and the displacement of the parts of body, and the way of its displacements in physical space [5] (=task space). Most of the studies on human motion consider the movement at the joint level [14–16], what is often called the *joint space* in robotics [11] Labanotation system segments a human body by *body signs* which specify limbs, joints, areas, and surface of the body [17]. Concerning the motion of limbs, though *space measurement signs* specify the degree of flexion of limbs, a movement does not be considered by only change of the degree of joints in Labanotation. Both cases in robotic and in Labanotation, data are referred as kinematics and are actually related [11–18].

One other interesting aspect is the contact with the environment and how the body generates force to interact with it and receives force from it [19, 20]. In Robotics it is called *external forces* [11]. Finally, the computational tools of Robotics [11] and Biomechanics [1] allow to also consider the inner forces acting on the body such as the joint torques and the muscle activity and muscle forces. In Labanotation, a notion of *dynamics* seems similar to these concepts of external and inner force in robotics. *Dynamics* in Labanotation links to the quality of the movement that varies according to flow of energy and force, different way of using time and space during a performance [8]. However, Labanotation does not measure *external force* which can affect human motion, nor *inner force*. Labanotation categorizes different accentuation, intensity or muscular tension observable during the execution of a movement by s*trength measurement signs* (*dynamics signs*) [5]. In Laban Movement Analysis (LMA), a concept of *Effort*, which is one of the four components of LMA, focuses on clarifying a motion dynamics [21]. Laban Movement Analysis is a method to analyse, observe and explore human motions, and it differs from Labanotation (Kinetography Laban) which aims to translate movement process taking place in four dimensions (three dimensional spaces and time) into signs written in two-dimension [18]. *Effort* analyses a human dynamic and its relations with personality or psychological reactions such as emotion, feeling, and thinking in a descriptive and indicative way by using a visual support, but *Effort* does not take account into how muscular forces generate such dynamics.

In this study we consider the dynamics in the general form of the equation of motion, when the word refers to the Labanotation *dynamics* it is explicitly mentioned, since this term in the Labanotation refers to the left hand term of Eq. (1) and eventually also includes $\boldsymbol{G_b}$ and $\boldsymbol{G_c}$ according to [8].

3 Feature Vector Analysis Using Principal Component Analysis (PCA)

In this section we present a framework to classify time series of data: here motion data using what is called a feature vector and introduced in Sect. 3.1. The feature vectors are then analyzed using principal component analysis (PCA) to obtain clouds of points. The points' proximity indicates resemblance in the motion data, and can form clusters of similar motion data as developed in Sect. 3.2. These clusters can then be used to analyze furthermore the data, develop recognition algorithms and so on as presented in Sect. 3.3.

3.1 Construction of the Feature Vector of a Data

A feature vector is a vector that contains characteristics that could quantify various data $\mathbf{q}_i$: kinematics, dynamics, in the joint space, in the task space, contact forces, muscle activity… It is obtained by computing the auto-correlation matrix of the considered data $\mathbf{q}_i$ [22]. The auto-correlation matrix $\mathbf{O}_i(l)$ of a given data-set is given by Eq. (2). l is a time constant difference, here we set $l = 2$, because we can reflect the information of the data change (velocity for example). Then we arrange the elements of $\mathbf{O}_i(l)$ into a single column vector. The result is called the feature vector for the given motion data $\mathbf{q}_i$ for an iteration i. For each motion repetition or each new motion i the feature vector is computed.

$$\mathbf{O}_i(l) = \frac{1}{T_i - 2} \sum_{k=l+1}^{T_i} \mathbf{q}_i[k]\mathbf{q}_i^{\mathrm{T}}[k-l] \quad (2)$$

3.2 Principal Component Analysis of the Feature Vector of a Data

Principal Component Analysis (PCA) of the obtained feature vectors provides information of the clustering properties of the data. A training data-set is created with a few exemplars of data. Consequently, it gives information on the possibility to discriminate a data-set from another data-set [22]. Applied to motion recognition, it means that it gives information on the differences and resemblances of different motion data-set; it allows discriminating between several motions, for which the algorithm was trained. This algorithm functions as an unsupervised learning algorithm, since data as data accumulates the PC space varies and clusters appear. Depending on the resemblances, points create clusters of various shapes, in the space of principal components, which are dense or scattered. It is also possible to find whether a motion belongs to the training data or not, and perform incremental learning.

Often the three dimensional space of the first three principal components is used because it allows a visualization of the dataset. The two dimensional principal component (PC) space can also be used if the cluster structure is clear enough using only the first two components. The shape of a cluster highlights data-set with similarities, while scattered points represent data-set with little similarity to each other.

3.3 Motion Classification Algorithm

The proposed recognition algorithm is based on the clustering in the PC space of the feature vectors [23]. In the feature vector PC space, we calculated a feature value using the center point of each cluster and the approximate straight line by the least-squares method of each cluster as shown in Fig. 1. It thus provides the center and the approximate shape of the cluster. When a test data needs to be compared with training dataset, it is projected into the PC space and the distance from each cluster center d_{ij} as well as the distance to the linear regression h_{ij} are calculated. The weighted sum of d_{ij} and h_{ij} gives the feature value S_{ij} which is used to determine whether the test data belongs to one of the existing clusters (if so which one) or not. The smallest S_{ij} allows to conclude that the test-data belongs to cluster Motion j. If it is not small enough then one can conclude that it belongs to none of the existing cluster and a new cluster can be created. The process is described in Fig. 2, where the test data belongs to the cluster “Motion 4”.

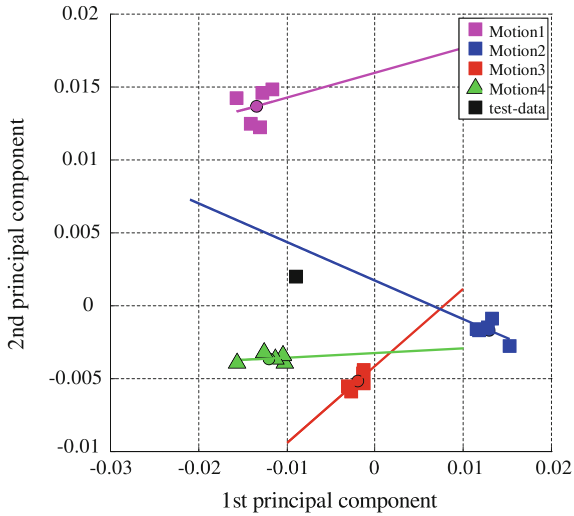


Fig. 1 Feature vector in the PC space forming clusters of similar motions for four examplar of motion data (*color of markers*) repeated each five times (*number of markers of each color*). And barycenters (*round marker*) and linear approximations (*straight line*) for each cluster

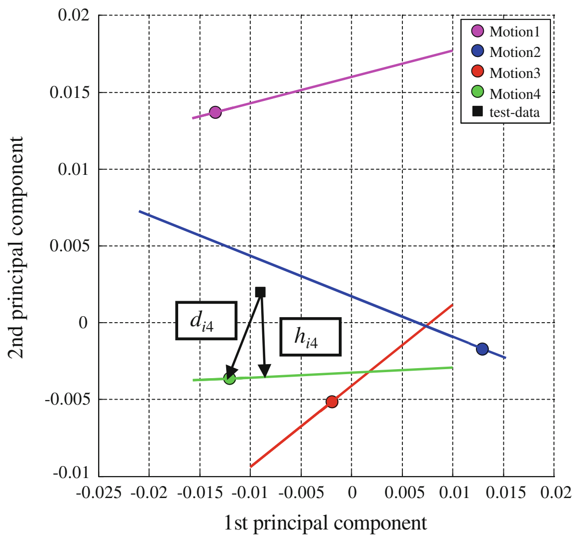


Fig. 2 Concept of the proposed recognition algorithm using the barycenter and the linear approximation for each clusters in the PC space of the feature vectors. The distance from each cluster to the test-data provides a metric: feature value that allows to associate the data with a cluster (in that case Motion 4), or without any and is the base of our recognition algorithm

The weights of the weighted sum can be chosen to give equal weight to the distance from the straight line and the center, or to favor one over the other depending on the shape of the cluster. For example a rather spherical cluster may give more weight to the distance to the center, while an oblong cluster may require more weight on the distance to the straight line. Using the training data to define the shape of the cluster can be used to generate automatically appropriate weights. These weights need to be recalculated any time a new test-data is inserted in the dataset since the addition of a new data point may change the global shape and size of the clusters.

4 Examples of Motion Classification and Motion Recognition

The feature vector can be obtained with any kind of data for any type of motion. It can be used to classify motions and there resemblance, it can be used to classify inner state that are embedded in a motion: moods and emotions. In fact it has been shown in the literature that motion carry more information than just the simple kinematics and dynamics, and these data relates to human inner states [24, 25]. In the following section we show some examples of utilization of our algorithm and the potential it can offer for automatic notations through quantification.

4.1 Using Kinematics Data in the Task Space: Gait

Gait is one of the most common task in daily life, and it is also the most important locomotion characteristics of human and humanoids. Gait is rich in information and it was shown extensively in our previous work that biometrics from gait is possible [23], as well as emotion recognition using the torso motion and the head inclination, with a simple similarity index [13]. Using the same dataset of 4 candidates and 5 repetitions of each emotional gait, which is extensively described in [13], the individual emotion recognition rate reaches 90 % in average. For further detail in the protocol and the results please refer to our works [13, 23] (Fig. 3).

4.2 Using Contact Force Data

The dynamics of the motion also contains rich information. Each motion has a dynamic signature that is characteristic with the way one interact with the environment and according to (1) it is possible to link the external contact forces and the motion directly. We propose to use the contact force information measured by force

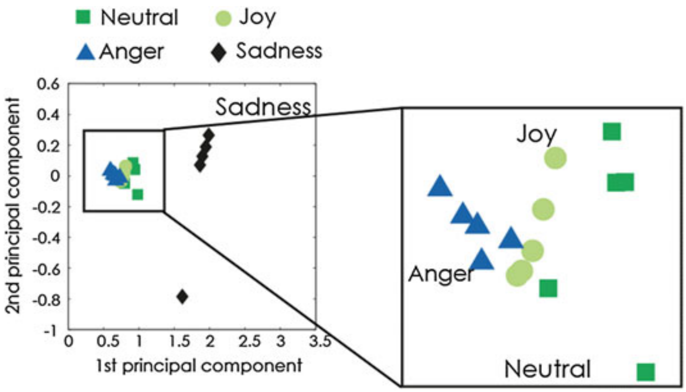


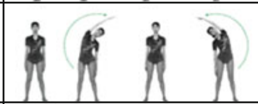
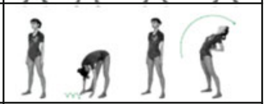
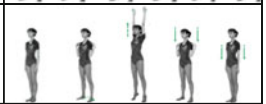
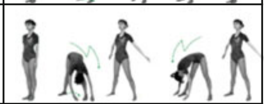
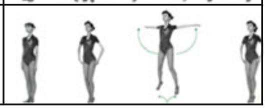
Fig. 3 Feature vector in the PC space for the four basic emotions in 5 gait trial for a given individual [13]. "Sadness" strikes out the rest of the emotions. The close up on the right shows that when sadness is not considered, each emotion also appears in a cluster, allowing for emotion recognition in gait

plates or insole sensors for example to classify motions with the algorithm proposed in Sect. 3. In our preliminary works [19, 24] we proved that it was possible to recognize motions using only the contact force measurement regardless of the person characteristics using a laboratory grade force-plate that provided the 3 forces and the 3 moment of forces. We now propose to use Nintendo wii balance boards that can only measure the vertical force and the moment of force in the horizontal plane. The algorithm is tested experimentally on the data of 5 subjects for 7 types of exercises motion.

4.2.1 Experimental Protocol

Figure 4 presents the type of motions that were performed on top of two Nintendo wii balance boards. This set of seven prescribed motions was chosen among a Japanese daily television exercise program (Radio Exercise). Five male candidates (mean age 24) were shown a video of the motions they were asked to perform, prior to the experiment. During the experiments, the same video was shown so that candidates can synchronize with the video. Each motion was repeated five times to insure enough training data and enough test data, in addition to the fact that a motion sequence may consist in repeating a few times the same motion. We measured the contact force information for each of the chosen seven sequences of motions noted M1–M7 as shown in Fig. 4.

Fig. 4 The seven different types of movements recorded on top of the forceplate for five male candidates. Each motion was repeated five times, to provide sufficient training data and test data

M1	Arm circles	
M2	Side-bending	
M3	Front bending	
M4	Waist rotations	
M5	Legs and arms	
M6	Touch their foot	
M7	Small jumps	

4.2.2 Results and Discussion

The recorded motions are automatically segmented [26]. The feature vectors were calculated with the vertical forces measured by the Nintendo wii balance board and the moment of force around the horizontal plane. In total 6 components. The classification results are given in Fig. 7. We use the exclusion method proposed in [23]. This suggests that with trained subject the repeatability would in-crease and that recognition would be much easier. With un-trained subject the size of the clusters increases. This can also be used as a quantification of training performance or motion repeatability.

Figure 5 show the PC space obtained for all the candidates and all the motions. Clear clusters appear in the PC space, while some data are more confused. The exclusion method automatically removes the data that are too far to narrow the search space of possible motions. The total average recognition rate reaches 75 %, as shown in Table 1. The confusion matrix given in Table 2, also shows that motions that have high similarities are more difficult to discriminate from each other, such as M3 and M6 that only differ from the direction of the bending. This suggests that some motions have poor repeatability: poor execution of the motion by the candidate for example. For example if one candidate moved his arm faster or slower, or with less amplitude in one motion, it is plotted at a point that is far from the motion's cluster. This suggests that with trained subject the repeatability would

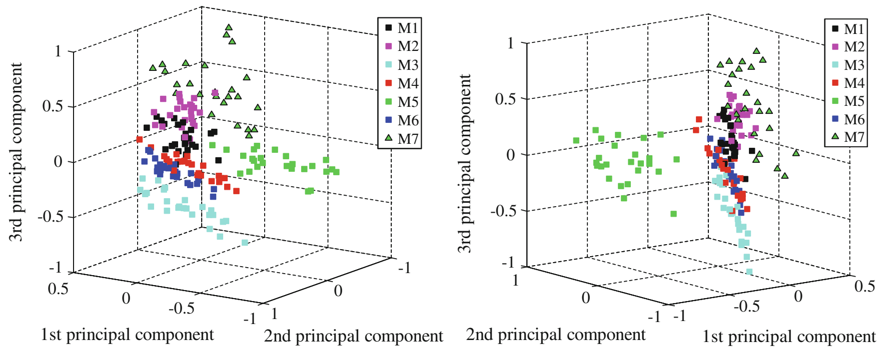


Fig. 5 Feature vector in the PC space for the seven motions repeated each five times, by five candidates. Both graphs represent the same results with a different view point

Table 1 Summary of the successful recognition rate [%] for each motion using the proposed exclusion method

M1	M2	M3	M4	M5	M6	M7
76	72	68	76	100	72	60

The average recognition rate is 75 %

Table 2 Summary of the confusion matrix (M1–M7) using the proposed exclusion method

	M1	M2	M3	M4	M5	M6	M7
M1	0.76	0.24	0	0.08	0	0.04	0
M2	0.08	0.72	0	0	0	0	0.32
M3	0	0	0.68	0.04	0	0.2	0
M4	0.08	0	0	0.76	0	0.04	0.08
M5	0	0	0	0	1	0	0
M6	0.04	0	0.32	0.12	0	0.72	0
M7	0.04	0.04	0	0	0	0	0.6

increase and that recognition would be much easier. With untrained subject the size of the clusters increases. This can also be used as a quantification of training performance or motion repeatability in terms of dynamics.

5 Application for Emotion and Mood Classification in Motion

5.1 Using Contact Forces

Using a similar experimental protocol as described in Sect. 4.2.1, candidates were asked to walk on top of the force plates and to change their mood while walking. Each data was taken 5 times. The obtained PC space is shown in Fig. 6. A clear

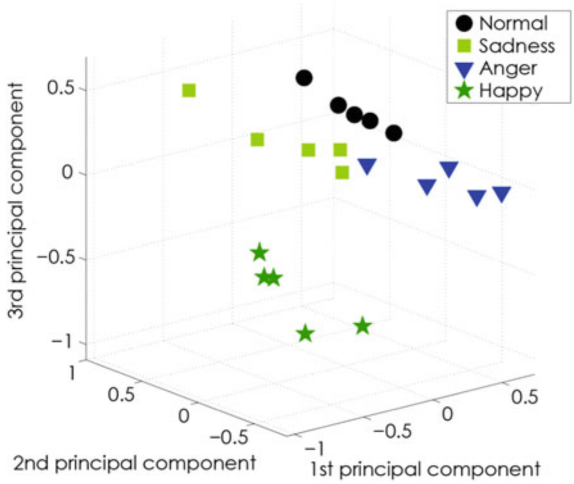


Fig. 6 PC space obtained from the contact force information while walking with different moods: normal, sadness, anger, happy. Each form a cluster charateristic showing that emotions can be recognized through the way one interact with the environment dynamically

cluster structure appears and the average recognition rate for all of the four emotions reaches 80 %. These results suggest that the contact forces with the environment and thus the dynamics contain also emotional information. This allows to quantify and classify dynamics data and thus will likely enable automatic classification and annotation of the *dynamics* in the Labanotation system and of the *Effort* in Laban Movement Analysis.

5.2 Using Kinematics and Muscle Contractions

As explained in Sect. 2, in Labanotation the dynamics describes a wide range of variables and among them the muscular intensity plays an important role. Muscular information relates with the dynamics through the joint torques. Moods show a variety of physiological manifestations that can be measured with a diverse array of techniques. The physiological activities such as brain waves, skin responses, heart responses and muscle responses, reflect change of a mental condition. The information on the mood included in behavior is classified into nonverbal information, and is also included in behavior without necessarily being based on the intention of an agent [27]. With surface electromyography (EMG) it is possible to record the muscular activity of surface muscles and associate it with motion and emotion. In this section we present our findings regarding the relationship between shoulder and upper-arm muscular contractions and mood variations. Identification of emotion and mood using EMG information has been done with a variety of methods until now [28–31]. Most are based on use of facial muscles. Walking is one of the key actions when identifying the information on humans. It is known that human walking includes information that is specific to the individual and be affected by

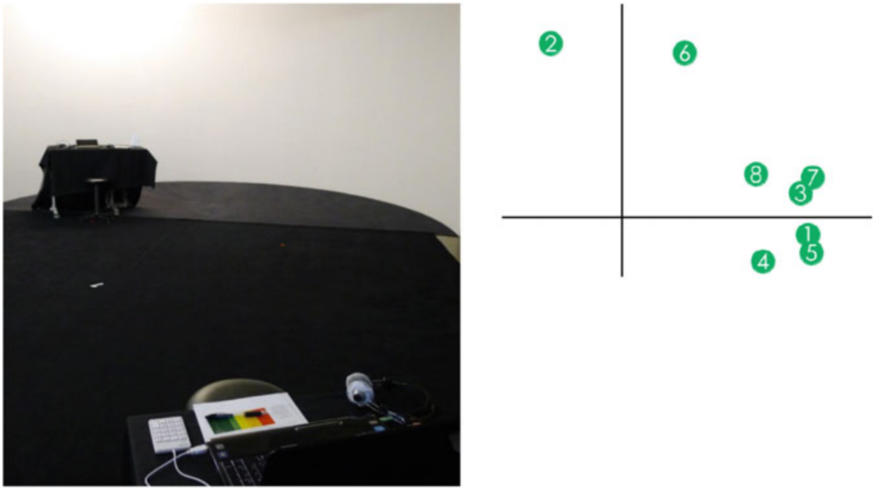


Fig. 7 Experimental setup for the mood changing experiment. *Left* The experimental setup at College de France with the two working stations and the free space for walking in between. *Right* Principal component analysis of the mood scores of the psychological TDMS test at the different stages of the experiments. The number in the *green circle* corresponds to the number of occurrence. It shows that the mood changes with time during the experiment as expected

mood [13, 24, 25]. That is, it is thought that the EMG analysis of walking is effective in the identification of human mood. In this work, we made a subject walk in various mood states and answer psychological tests that measure the mood. We use two types of tasks: music listening (affective value scale of music (AVSM) and multiple mood scale (MMS) [32]); and numerical calculation for evoking different moods. It is expected that numerical calculation evokes either a more aggressive or frustrated mood upon failure, or self-satisfaction when successful, than listening to music. Figure 7 shows the experimental protocol of our experiment. In foreground space the task is listening to calm music (odd numbers in the green circles). In the background space the subject's task which is either listening to music (uplifting and depressive) or doing the numerical calculation (even numbers in the green circles). Inserting a task of listening to calm music between each other tasks allow to stabilize the moods as can be seen from Fig. 7-right.

Statistical features of EMG signals are calculated using Fast Fourier Transform and Principal Component Analysis. These statistical features are related with psychological test scores obtained from a two-dimensional Mood Scale (TDMS) [33], using regression analysis.

The two types of tasks are performed on two laptop computers. Moreover, the tasks are done by turns at two distant spaces in the room, and the subjects walk back and forth between them. After completion of each task, the subjects fill in questionnaires about their mood at that time, and then the subjects walk to the other space. The TDMS test consists of eight questions on a 6-point Likert scale, and can quantify the state of mind at the time of measurement. The result of TDMS is expressed as a score of "pleasure" and "arousal" (from −20 to 20). In other words, high pleasure indicates a comfortable and positive state, high arousal indicates an excited and active state. The test is filled in less than 1 min, thus it is suitable for temporal observation of mood changes. The subjects repeated that work 8 times. Subjects were not informed that the motion for interest in our study was walking to insure walking as naturally as possible.

Surface EMGs of the biceps, triceps, middle deltoids, and upper trapezius were recorded. The mean power frequency MPF obtained by Fast Fourier Transform (FFT) of the EMG data were generated for each muscles for the PCA.

$$\mathrm{MPF} = \frac{\int_0^\infty f P(f) \mathrm{d}f}{\int_0^\infty P(f) \mathrm{d}f} \tag{3}$$

where P is the power of the signal, and f is its frequency.

The PCA was performed for each frequency range over all the muscles. Finally a multiple linear regression analysis (MRA) was performed on the principal component score as an explanatory variable and the amount of variation in the TDMS pleasure score which is the difference between each pleasure score and the previous one as a response variable. By considering the score not using an absolute value but using a relative value, we can take into consideration the individual difference of the reaction to a stimulus.

5.2.1 Experimental Results and Discussion

As a result of the regression analysis as shown in Fig. 8, we can confirm a statistically significant positive correlation between the muscle activity and the arousal level (p < 0.001) with R = 0.59. This correlation is promising in using muscle information to predict moods. Moreover muscle contraction are related to more observable variable such as joint stiffness, allowing for quantification of mood/emotion through joint stiffness.

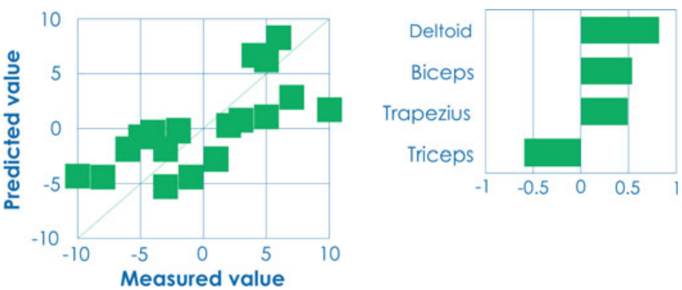


Fig. 8 Experimental results. *Left* Mean power frequency of EMGs of muscles around the shoulders and psychological TDMS test scores of "pleasure" correlates. EMGs can be used to predict the mood state. *Right* Loading factors of the different muscles involved

6 Conclusion and Perspectives

This chapter emphasizes on considering not only the kinematics information of a motion but also the dynamics information, and in particular the contact forces with the environment when analyzing human motion or simulating motions for animation or humanoid robotics. Indeed, the dynamics play an important role in robotics to control the stability and generate smooth and safe motions. In biomechanics it allows to understand the interaction of the human body with the environment and in Labanotation, dynamic is important criteria to observe and describe the way and the quality of movement and performance. The theory of *Effort* in Laban Movement Analysis treats exclusively a study of dynamic and explores its relation with psychological state with using a rigorous observation grid [21]. Using a quantification, a classification and a systematic analysis of the dynamics as presented here will allow to generating systematic dynamics annotations, both as understood in robotics [11] as well as in the Labanotation [8]. This systematic notations will help in bridging the gap of terminology of both communities and in providing a general framework and understanding that will benefit both communities. Our promising results in using contact forces, muscle activity and further joint stiffness to understand mood and emotions are also crucial to generate motions that can convey feelings, and understand these motions when performed by humans.

If some of these results are at a preliminary stage, we are confident that they will lead to outstanding outcome in human motion science and further to humanoid motion generation.

References

1. A. Torezen, *Human Body Dynamics: Classical Mechanics and Human Movement* (Springer, Berlin, 2000)
2. M.Z. Patoli et al., Real time online motion capture for entertainment applications, in *The 3rd IEEE International Conference on Digital Game and Intelligent Toy Enhanced Learning DIGITEL* (2010), pp. 139–145
3. A. Fern'ndez-Baena et al., Biomechanical validation of upper-body and lower-body joint movements of kinect motion capture data for rehabilitation treatments, in *Proceedings of 4th International Conference on INCoS* (2012), pp. 656–661
4. F. Loula et al., Recognizing people from their movement. J. Exp. Psychol. Human Percept. Perform. **31**(1):210–220 (2005)
5. A. Hutchinson Guest, *Choreographics: A Comparison of Dance Notation Systems from the Fifteenth Century to the Present* (Routledge, 1989)
6. W. Choensawat, M. Nakamura, K. Hachimura, GenLaban: a tool for generating Labanotation from motion capture data. Multimedia Tools Appl. 1–24 (2014)
7. A. Soga, B. Umino, T. Yasuda, S. Yokoi, Automatic composition and simulation system for ballet sequences. Vis. Comput. **23**(5), 309–316 (2007)
8. A. Hutchinson Guest, *Labanotation: The System of Analyzing and Recording Movement*, 4th edn. (Routledge, 2005)

9. K. Yoshida, D. Nenchev, M. Uchiyama, Moving base robotics and reaction management control, in *Proceedings of the 7th International Symposium of Robotics Research* (1995), pp. 100–109
10. Y. Fujimoto, S. Obata, A. Kawamura, Robust biped walking with active interaction control between foot and ground, in *Proceedings of the IEEE International Conference on Robotics and Automation* (1998),pp. 2030–2035
11. W. Khalil, E. Dombre, *Modeling, Identification and Control of Robots* (Hermès Penton, London, 2002)
12. A, Safonova, J.K. Hodgins, N.S. Pollard, Synthesizing physically realistic human motion in low-dimensional, behavior-specific spaces, in *ACM Transactions on Graphics*, vol. 23, no. 3 (ACM, 2004), pp. 514–521
13. G. Venture, H. Kadone, T. Zhang, J. Grezes, A. Berthoz, H. Hicheur, Recognizing emotions conveyed by Human Gait. Int. J. Social Robot. **6**(4), 621–632 (2014)
14. W. Takano et al., Humanoid robot's autonomous acquisition of proto-symbols through motion segmentation, in *IEEE International Conference on Humanoid Robots* (2006), pp. 425–431
15. D. Kulic et al., Online segmentation and clustering from continuous observation of whole body motions. IEEE Trans. Rob. **25**(5), 1158–1166 (2009)
16. F. Ofli, R. Chaudhry, G. Kurillo, R. Vidal, R. Bajcsy, Sequence of the most informative joints (smij): a new representation for human skeletal action recognition. J. Vis. Commun. Image Represent. **25**(1), 24–38 (2014)
17. A. Hutchinson Guest, *Dance Notation: The process of recording Movement on Paper* (Dance Horizons, New York, 1984)
18. J.P. Scholz, G. Schöner, The uncontrolled manifold concept: identifying control variables for a functional task. Exp. Brain Res. **126**(3), 289–306 (1999)
19. C. Yang, G. Ganesh, S. Haddadin, S. Parusel, A. Albu-Schaeffer, E. Burdet, Human-like adaptation of force and impedance in stable and unstable interactions. IEEE Trans. Robot. **27** (5), 918–930 (2011)
20. G. Venture, C. Hamon, Motion recognition from contact forces information and identification of the human body dynamics, in *IEEE/RAS-EMBS Proceedings of International Conference on Biomedical Robotics and Biomechatronics* (Tokyo, Japan, 2010), pp. 295–300
21. R. Laban, *The Mastery of Movement*. 4th edn. ed. by L. Ullmann. (MacDonald and Evans, London, 1980). (First published as The Mastery of Movement on the Stage, 1950
22. H. Kadone, Y. Nakamura, Symbolic memory of motion patterns by an associative memory dynamics with self-organizing non-monotonicity. Curr. Biol. **18**(8), 300–329 (2008)
23. T. Zhang, G. Venture, Individual recognition from gait using feature value method. Cybern. Inf. Technol. **12**(3), 86–95 (2012)
24. B. de Gelder, Towards the neurobiology of emotional body language. Nat Rev Neuroscience **7**, 242–249 (2006)
25. C. Roether, L. Omlor, A. Christensen, M. Giese, Critical features for the perception of emotion from gait. J. Vis. **9**(6), 1–32 (2009)
26. T. Yabuki, G. Venture, Human motion classification and recognition using only contact force, in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems* (Hamburg, Germany, 2015). (submitted)
27. P.M.A. Desmet, Measuring emotions, development and application of an instrument to measure emotional responses to products. in *Funology: From Usability to Enjoyment* eds. by M.A. Blythe, A.F. Monk, K. Overbeeke, P.C. Wright (Kluwer Academic Publishers, 2003), pp. 111–123
28. X. Zhu, Emotion recognition of EMG based on BP neural network, in *Proceedings of the Second International Symposium on Networking and Network Security* (2010), pp. 227–229
29. M. Murugappan, Electromyogram signal based human emotion classification using KNN and LDA, in *IEEE International Conference on System Engineering and Technology* (2011), pp. 106–110
30. T. Partala, V. Surakka, T. Vanhala, Real-time estimation of emotional experiences from facial expressions. Interact. Comput. **18**, 208–226 (2006)

31. A. Gruebler, K. Suzuki, Measurement of distal EMG Signals using a wearable device for reading facial expressions, in *Engineering for Medicine and Biology Conference (EMBC)* (2010), pp. 4594–4597
32. T. Taniguchi, Construction of an affective value scale of music and examination of relations between the scale and a multiple mood scale. Jpn. J. Psychol. **65**(6), 463–470 (1995)
33. Y. Sakairi, H. Tokuda, M. Kawahara, T. Tanigi, H. Soya, Development of the two dimension mood scale for measuring psychological arousal level and hedonic tone. Bull. Inst. Health Sport Sci. Univ. Tsukuba **26**:27–36 (2003)
34. T. Yabuki, G. Venture, Motion recognition from contact force measurement, in *Proceedings of the IEEE International Conference Engineering in Medicine and Biology* (Osaka, Japan, 2013), pp. 7245–7248

The Effect of Gravity on Perceived Affective Quality of Robot Movement

Suzanne Weller, Joost Broekens and Gabriel A.D. Lopes

Abstract Non-verbal communication, in particular emotions and social signals, has the potential to improve interaction between humans and robots. Body movement style is known for influencing the affective interpretation of a movement in humans. In this paper the effect of gravity on perceived affective quality of robot movement is investigated. Simulations of a robot arm executing various daily tasks were created. Each task is executed under three different virtual gravity conditions: positive (downward directed force), negative (upward directed force) and no gravity. In a user study participants rated videos of the movement of the robot arm in terms of its emotional content. The robotic arm performed ten different tasks. Two response tools were used for the participants to rate the videos: the AffectButton and the Self-Assessment Manikin. Results show that there was a residual significant effect of the virtual gravity variable on the AffectButton. Moreover, there was a large significant effect of task on the ratings of both the AffectButton and the Self-Assessment Manikin. This indicates that gravity has a small, but measurable effect on the perceived emotional content of even a simple, rather disembodied, robot movement.

S. Weller
BioRobotics Department, Delft University of Technology,
Delft, The Netherlands
e-mail: smcweller@gmail.com

J. Broekens
Interactive Intelligence Group, Delft University of Technology,
Delft, The Netherlands
e-mail: D.J.Broekens@tudelft.nl

G.A.D. Lopes (✉)
Delft Center for Systems and Control, Delft University of Technology,
Delft, The Netherlands
e-mail: G.A.DelgadoLopes@tudelft.nl

J.-P. Laumond and N. Abe (eds.), *Dance Notations and Robot Motion*,
Springer Tracts in Advanced Robotics 111,
DOI 10.1007/978-3-319-25739-6_18

1 Introduction

It is widely accepted that robots are entering our societies and will be ubiquitous in the near future. Robots that exist today already resemble humans [1]. Creating robots that move as humans is, however, a challenging but still distant goal. Replicating human agility requires new types of actuators and power sources. Imitating human's capability for nuanced motion and expressive gesture requires novel control algorithms. Human-like motion also has the potential to mitigate the well known Uncanny Valley [2]. Non-anthropomorphic robots can also benefit from the knowledge of human movement. The first reason for considering non-humanoid expressive robots, is that robots that are specially designed for certain tasks, do not have to be limited to human-like morphology. The second reason is that of feasibility in terms of market economy. At this moment, humanoid robots are developed in research laboratories and are not yet on the consumer market. Simpler non-humanoid robots hold cost-effective designs, allowing for large-scale replication [3]. This paper is a contribution to the field of bodily emotion expression in non-anthropomorphic robotics by analyzing the human perceived effect of virtual gravity on a collection of simulated tasks generated by a class of control laws.

1.1 Related Research

One of the key challenges in bodily emotion expression lies on the multitude of theories that attempt to define emotions and how they can be identified and structured. When it comes to non-verbal bodily expression, different concepts have been proposed on which variables are changed when expressing certain emotions. On the one hand researchers have looked at the different body parts that are used, for example [4–9]. On the other hand studies have investigated the characteristics of the motions of these body parts but also of whole body movement, for example [5, 10–15]. The next development step involved the transition of these ideas into the expression of emotions by animated characters or robots. Again, different approaches exist here. Some studies have focused on mimicking key poses from actors: [16–20], others have used movement characteristics from human motion studies to control virtual characters or robots: [21–24].

There is no complete and agreed view upon a standard for emotion expression in robots. As such there is a need for validated principles that can be used to generate emotions in the bodily motions of robots. Today researches have approached this problem by investigating the use of different variables (e.g. position, velocity, extensity [23]). This paper follows a similar approach, and focuses on the validation of a new parameter: virtual gravity. The hypothesis is that a movement generated with an upright open posture and 'high energy' used to overcome gravity reflect a high level of dominance, pleasure and arousal, while a movement with a low closed posture is perceived in the opposite way. In nature, some land animals, such as grizzly bears, often use upright postures to display dominance. We see this as corroborating our hypothesis. Cuttle fish use flashing patterns of colour in their bodies to signal dominance. This is probably a consequence of their water environment where gravity does not affect the dynamics of sea creatures (they are mostly naturally buoyant). Thus, raising tentacles is probably not perceived as a dominant behaviour in a water environment by other sea creatures. Humans live in an environment strongly affected by gravity. As such we hypothesise that the apparent effect of gravity on the body can lead to the perception of different emotional content.

2 Model-Based Control

In this section we introduce a mathematical model to test our hypothesis: virtual gravity is a parameter that influences the perceived emotion of a generated robotic movement. The model consists of two parts: first the physics-based model of the robot, and second, the mathematical model of the controller.

2.1 Model and Controller Design

In this paper we use a model-based control approach. We consider a full dynamical model of the robotic platform to enable changing the effect of gravity. This effect is implemented by changing the gain on the gravity compensation term. The proposed control law is shown in (1)

$$u = -J^T\left(K_p(x - x_d)\right) - K_d\dot{q} + (1 - \delta)G(q), \tag{1}$$

applied to the standard mechanical model of a robotic manipulator [25], described by

$$M(q)\ddot{q} + C(q, \dot{q})\dot{q} + G(q) = u. \tag{2}$$

The control law u is constructed using three elements. The first element $J^T\left(K_p(x - x_d)\right)$ is the task controller part (in practice it consists of projecting a spring-based virtual force in the workspace into the joint coordinates). The second element, $K_d\dot{q}$, realises energy dissipation. The third element modulates gravity compensation. Applying the control law (1) into (2) results in:

$$M(q)\ddot{q} + C(q, \dot{q})\dot{q} + \delta G(q) = -J^T\left(K_p(x - x_d)\right) - K_d\dot{q}.$$

Knowing that one can write $\delta G(q) = \delta g \bar{G}(q)$, where g is the standard acceleration due to gravity, and $\bar{G}(q)$ is only a function of the kinematics of the robot, then the parameter δ represents the number of virtual g-forces acting on the robot. For example,if $\delta = 0$ the robot is not affected by gravity, $\delta = 1$ results in normal gravity, and $\delta = -1$ results in the effect of a reversed gravity vector. With this control law, the task of the end effector is achieved by the first term, dissipation on the entire arm by the second, and the third term in effect uses the extra degrees of freedom to react to the δ effect of virtual gravity. By changing δ one influences all body parts of the robot between the base and the end effector. A task trajectory can then be executed under different gravity circumstances.

2.2 Implementation

The goal is to create simulations where the effect of a changing virtual gravity vector on the robotic bodily movement is clearly visible. A robotic arm was chosen to execute various tasks. While executing a task, the changing virtual gravity affects the way the body performs the task without interfering with the task itself, i.e. the movement of the manipulator is the same. The full dynamical model of the robotic arm was implemented in Matlab using Lagrangian mechanics with joint angles as the generalized coordinates. It was decided to simulate an industrial looking robotic arm, with 10 degrees of freedom. The model of the arm can be seen in Fig. 1.

The evolution of the joint angles for a given task trajectory described by expression (1) was solved using a standard Newton-Euler integration method with variable sampling rate.

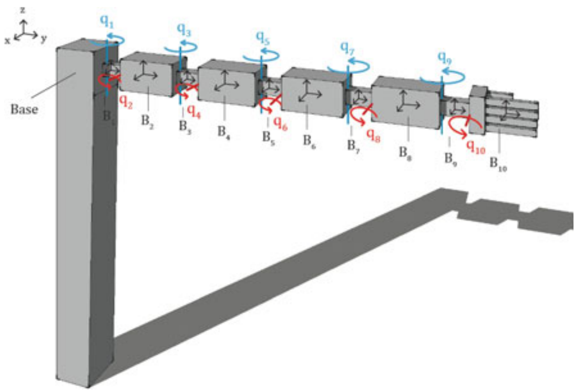


Fig. 1 Model of the simulated arm

3 Methodology

3.1 Simulations

Ten tasks were chosen for the robotic arm. While executing these tasks, the changing virtual gravity vector affects the way the arm performs the task. The next section describes the experimental design choices in more detail, and the steps taken to create the simulation videos.

1. Arms: Two robotic arms are used to execute each task, in order to verify that the responses from the user study are independent of the visual appearance of the arm. A third arm was designed only to use for the start-up test of the experiment. The arms were designed to have industrial features. They purposely have kinematic differences from a human arm. For some tasks, it was convenient to add an human-like hand. The first arm is square-shaped with open centres. The second arm is similar, but has round shapes. The third arm is also round-shaped, but all bodies are of an equal size. Figure 2 illustrates the final design of the three arms.
   All arms are composed of 10 rigid bodies, connected by 10 rotating joints, resulting in 10 actuated degrees of freedom. This over-dimensioned number of joints was chosen to give the arms enough freedom to simultaneously perform the end-effector task in the workspace and result in different body postures for different gravity levels.
2. Tasks: Daily tasks were given to the robotic arm to execute. Ten tasks were designed to be used in the main part of the experiment: closing a book, opening a door, opening a drawer, giving an object, pointing at an object, pushing an object towards a person, replacing an object, stirring a bowl, switching on the light and writing on a blackboard. One extra task, opening a box, was designed and used for the start-up test of the experiment. Different tasks were designed in order to make sure that the responses from the user study would be independent of the type of task. The tasks were selected from the perspective that they do not have an initial emotional content. The selection of the tasks was based on the fact that some include interaction with an object, some include interaction with a person, and some include neither object nor person interaction. Since the focus

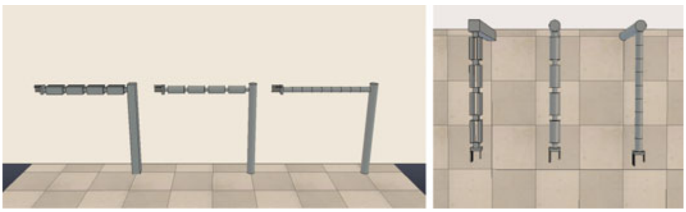


Fig. 2 Geometry of three non-human like robotic arms utilized in the user studies

of this experiment was on the changing effect of virtual gravity in the entire arm movements, all tasks were designed to include movements of the entire arm.

3. Virtual gravity: Three different levels of the virtual gravity vector were explored. Firstly, twice the normal gravity level, secondly no gravity, and thirdly twice the normal level in the reversed direction. These magnitude levels were chosen in order to obtain viewable differences in the simulations.
4. Visualisation environment: The software *Virtual Reality Education Pathfinder* or Vrep was used for the visualisation of the dynamic movement of the robotic arm. Together with the calculated dynamics of the arm, a natural looking scene was created around the robot. This was not only done to make the scene look more realistic, but also to give an indication of the size of the arm. The Matlab model of the arm was recreated in Vrep, using rigid bodies and joints. Six different videos were made for each task, using the combination of two arms and three levels of gravity. All six videos were recorded from the same view angle.

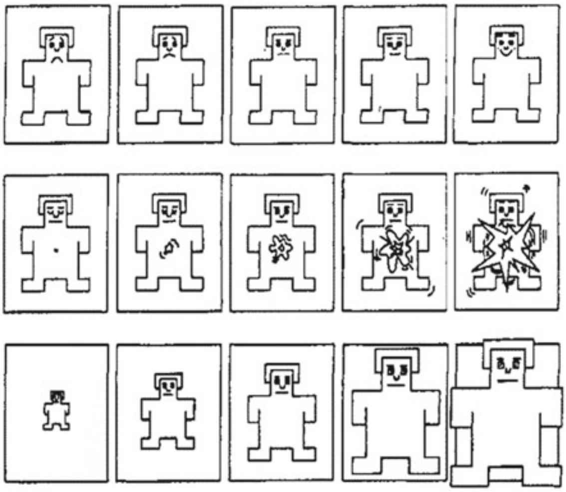
3.2 Experimental Setup

The statistical setup of the test is based on the main question of this research. The main goal of the user study is to see if participants can objectively identify emotional content in the movement of a robotic arm. In other words, it is *not* desired that they see the variations of one task as a result to the changing gravity, and then rate the emotional state while comparing these videos. The same holds for the morphology of the arm. Therefore, it was decided to introduce a “between subject” study for the gravity and the morphology of the arm. The third variable (task), was set up as a “within subject” variable. By implementing the two between subject variables: morphology (two options) and gravity (three options), six experimental groups were created. A website was constructed in order to present the experiment on-line.

1. Response tools: Two response tools were chosen to facilitate the participants to report the perceived emotional content of the robot movement in the simulation. In both tools graphical expressions were used to select an emotion. This is an advantage, since the participants do not have to use words to express an emotion.
   The first response tool is called the *AffectButton (AB)*, created by Broekens and Brinkman [26]. It is an on-line interactive button that one can control with a computer mouse. When moving across it, it changes it’s facial expression. When a certain facial expression is selected, the accompanied values for pleasure, arousal and dominance are saved. Figure 3 illustrates the AffectButton and several examples of affective states.

Fig. 3 The AffectButton with eight example expressions used to report on different levels of pleasure, arousal and dominance

Fig. 4 The self-assessment manikin used to report on different levels of pleasure, arousal and dominance



The second response tool is called the *Self-Assessment Manikin (SAM)*. This response tool was created by Bradley and Lang in 1994 [27]. It asks the participants for a direct score on pleasure, arousal and dominance. It uses a nine-point Likert scale accompanied with five supporting images. The Self-Assessment Manikin is illustrated in Fig. 4.

2. Questionnaire: The experiment was concluded with a short questionnaire. In total seven questions were asked concerning the participants gender, age, nationality and finally expert level in the following four categories: robotics, human movement analysis, acting and the interpretation and measurement of emotion.

4 Experimental Results

All effects are reported as significant at p < 0.05.

4.1 Main Analysis

To identify if gravity has an effect on the perception of emotion in the arm movements presented in the tasks, a multivariate mixed ANOVA of both the AB and SAM together is computed. Using Wilks's statistic, there was a near significant effect of gravity on the ratings of pleasure-arousal-dominance (PAD) for both the AB and SAM, $F(12, 528) = 1.630$, $p = 0.080$, $\eta^2_{partial} = 0.036$. Furthermore, there was no significant effect of morphology on the ratings of PAD for both the AB and SAM. In addition, there was no significant interaction effect between gravity and morphology on the ratings of PAD for both the AB and SAM. There was a significant effect of task on the ratings of PAD for the multivariate analysis, $F(54, 216) = 12.169$, $p < 0.05$, $\eta^2_{partial}$.

The next multivariate mixed ANOVA was executed only for the AB. Results show that there was a significant effect of gravity on the ratings of PAD for the AB, using Wilks's statistic, $F(6, 534) = 2.502$, $p = 0.021$, $\eta^2_{partial} = 0.027$. There was again a significant effect of task on the ratings of PAD for the AB, $F(27, 243) = 14.312$, $p < 0.05$, $\eta^2_{partial} = 0.614$.

Next, the same multivariate mixed ANOVA was executed only now for the SAM. Using Wilks's statistic, there was no significant effect of gravity on the ratings of PAD for SAM. Again, there was a significant effect of task on the ratings of PAD for the SAM, $F(27, 243) = 20.458$, $p < 0.05$, $\eta^2_{partial} = 0.694$.

Furthermore, when looking at the individual dependent parameters in a univariate ANOVA the following results are found. There was a near significant effect of gravity on the variable pleasure of the AB $F(2, 269) = 2.757$, $p = 0.065$, $\eta^2_{partial} = 0.020$. Also there was a significant effect of gravity on the variable dominance of the AB $F(2, 269) = 2.757$, $p = 0.046$, $\eta^2_{partial} = 0.023$. For the SAM, there was a significant effect of gravity on the variable pleasure $F(2, 269) = 3.523$, $p = 0.031$, $\eta^2_{partial} = 0.026$. On the other variables there was no significant effect.

The effect of task was also evaluated for the individual dependent parameters (e.g. AB pleasure, AB arousal, etc.). Mauchly's sphericity test pointed out that sphericity had been violated regarding five dependent variables. Only for the arousal of the AB the sphericity could be assumed. For the five other variables the degrees of freedom were corrected using the Greenhouse-Geisser estimates for sphericity (ε). Results show that for all six variables there was a significant effect of task on each individual variable, as illustrated in Table 1.

Table 1 Effect of task on dependent variables

Variable	Test	p	$\eta^2_{partial}$
AB pleasure	$F(8.248, 2421) = 35.902$	<0.001	0.118
AB arousal	$F(9.000, 2421) = 10.433$	<0.001	0.037
AB dominance	$F(8.301, 2421) = 13.796$	<0.001	0.049
SAM pleasure	$F(8.250, 2421) = 44.093$	<0.001	0.141
SAM arousal	$F(8.389, 2421) = 17.578$	<0.001	0.061
SAM dominance	$F(7.899, 2421) = 7.615$	<0.001	0.028

Since the significant effect of each task was present in all variables, further evaluation was done to evaluate the different effect of each task on the individual dependent variables.

The estimated marginal mean of each task was determined per dependent variable. These means were compared to the mean of each variable. Some of the tasks are more often significantly different from the mean of that variable than other tasks. The number of times the estimated marginal mean differed from the variable mean was counted. In Table 2 it can be seen that task 2, task 3, task 4 and task 5 show significant differences with the means in five out of the six dependent variables. These were the tasks of opening a door, opening a drawer, giving an object and pointing at an object. Task 8, stirring a bowl was rated significantly different for four out of the six dependent variables.

4.2 Secondary Analysis

The first part of the secondary analysis concerns the correlation between the two response tools. The results in relation to the differences in demography of the participants have been evaluated in the second part.

Table 2 Number of times of significant different results compared to the the variable's mean. The subscripts on the column headers mean pleasure, arousal, or dominance (e.g. AB_p represents AffectButton and pleasure)

Task	AB_p	AB_a	AB_d	SAM_p	SAM_a	SAM_d	Total
1		Yes			Yes		2
2	Yes	Yes	Yes	Yes	Yes		5
3	Yes	Yes	Yes	Yes	Yes		5
4	Yes	Yes	Yes	Yes	Yes		5
5	Yes		Yes	Yes	Yes	Yes	5
6						Yes	1
7					Yes	Yes	2
8	Yes		Yes	Yes		Yes	4
9							0
10		Yes					1

4.2.1 Correlation Between the AffectButton and the Self-assessment Maninkin

A Pearson's correlation test was executed for the three dependent variables pleasure, arousal and dominance. To do this the means of the different dependent variables were calculated for the ten tasks. These ten means were then compared between the AB and the SAM. There was a significant relation between the pleasure of the AB and the pleasure of the SAM, $r = 0.992$, $p\,(\text{two tailed}) < 0.001$. There was also a significant relation between the arousal of the AB and the arousal of the SAM, $r = 0.799$, $p\,(\text{two tailed}) = 0.006$. Finally, there was a significant relation between the dominance of the AB and the dominance of the SAM, $r = 0.729$, $p\,(\text{two tailed}) = 0.017$.

4.2.2 Demographic Information on the Participants

Gender

In total 275 persons participated in this experiment. 139 (50.5 %) men and 121 (44.0 %) women. Fifteen people (5.5 %) did not specify their gender. Most interesting result was that there was a small significant interaction effect of gravity and gender on the dominance of the AB, $F(2, 248) = 3.466$, $p = 0.033$, $\eta^2_{partial} = 0.027$.

Age

Most participants were between 19 and 40 years old. Nobody above 80 participated. Three people did not specify their age group. No interaction effect analyses were executed regarding the differences in age, since the distribution between the groups was not uniform.

Nationality

It was found that most participant (85.4 %) were from The Netherlands. No interaction effect analyses were executed regarding the differences in nationality, since the distribution between the groups was not uniform.

Expert Level

A participant was named an expert if for all four questions the response was never "None" (answer 1 out of 5). With this criteria 53 participants (19.3 %) were considered to be experts. Most interesting results show that there was no significant effects of gravity neither on the evaluation of the multivariate ANOVA's of the AB and SAM nor on any of the individual dependent variables.

5 Discussion and Recommendations

5.1 Gravity

It was seen that there was a significant effect of gravity measured on the AB. Moreover, in the univariate ANOVA there was a significant effect measured on the single variables AB dominance and SAM pleasure. However, it should be mentioned that in all cases this effect was very small.

The intermediate differences between the three gravity conditions were rather small. As a result the estimated marginal means were positioned very close together. We believe that this result arises from the lack of anthropomorphic elements in the arm we have used. In the human, the posture of the shoulders and inclination of the head give many clues to its emotional state. What the results show is that, without prior knowledge, it can be difficult for a human to perceive pleasure, arousal or dominance, in a non-human, non-animal like robot device. This suggests that a reductionist approach to modulating movement may be challenging. For future research it would be interesting to search for the minimum set of anthropomorphic-like elements that can generate the perception of emotional content without requiring a learning/training period by the human.

5.2 Task

It was seen that the within subject variable tasks gave large significant effects on both the AB and the SAM. For example the effect of a task on the AB ratings was 20 times larger than the effect of gravity. This can be explained as follows. Participants saw all tasks, hence there is a natural tendency to amplify the differences between the tasks. We postulate that the participants feel the need to 'look for' an affective difference between the perceived stimuli. A related explanation is that some tasks simply are perceived to be more positive (e.g., handing over a cup can be seen as polite or in service of) than others. Because the gravity factor was varied between subject, and task within subject, differences between tasks could be compared but differences between gravity setting could not. The large effect of task is thus a solid indication that we have made the right choice to study gravity as a between subject factor: if it had been a within subject factor participants would have also searched for a meaning and since the only thing they could rate was affect this would have inflated our effect size due to comparison effects.

5.3 Morphology

It can be stated that the effect of morphology was not present in this study. Apart from the very small significant interaction effect of task and morphology on the AB dominance, it can be concluded that the two arms morphologies were rated similarly. This means that the effect of gravity is not related to arm morphology, at least not to the morphologies we tested. For further research, regarding the movement of robot arms that are more or less similar when it comes to number of segments, degrees of freedom and size, this variable can be taken out of the experiment, and the use of only one type of arm would be sufficient. However, different robot bodies should be investigated. For example, robots with a more human-like body, or robots with more than an arm, as the effect of gravity could easily be larger in these cases.

5.4 Final Remarks

In this study, gravity as a parameter was singled out. It was the only variable that was tested in this experiment, while other variables that could influence the emotional perception were kept constant. In other studies on variables affecting the perception of the emotional content of robot movement, mostly multiple parameters were simultaneously tested. The study of Yamaguchi et al. [23] for example used position, speed and extensity. These studies argue that some emotions (anger and joy) could not be distinguished by only one parameter, in their case the velocity of the movement. However, Yamaguchi's research was different in that only four discrete emotions were directly generated, while in this research movements were evaluated by different levels of pleasure, arousal and dominance. With the use of the dimensional scale, more nuance is possible, which creates an opportunity for the nuanced effect of gravity to be measurable. Overall, we are convinced that distinguishing gravity with the used method was a correct way of testing one variable. However, the effect remains rather low. Therefore, future research could focus on implementing the gravity parameter, in association with a set of other variables, as for example velocities and accelerations, together with different control strategies.

One of the elements specific to this research is the deliberate use of a non-anthropomorphic robot. This was done to simplify the computational model, and was also based on research by Sawada [28] who showed that in humans it is possible to recognize emotions in arm movements. However, in this experiment the arm was attached to a fixed structure that was not affected by the changes of the virtual gravity. Mounting the arm on a different structure, possibly mobile, may lead to a better human interpretation.

6 Conclusion

In this paper we have concluded that gravity has an effect on the perceived affective quality of robot movement. We have shown this using a minimalistic setup in which a simulated disembodied robot arm was configured to do a set of service tasks (e.g., picking up a cup, opening a door, writing on a whiteboard, etc.). Participants could only see one gravity condition (e.g., positive, negative, or without) which influenced how the arm movement and posture was executed. Our experimental setting tested the effect of gravity in very strict conditions (i.e., between subject comparison, wide variety of tasks, two robot morphologies, and a disembodied non human-like arm). Therefore, we conclude that gravity can modulate the affective quality of robot movement, even though this effect was small.

References

1. K.F. MacDorman, H. Ishiguro, The uncanny advantage of using androids in cognitive and social science research. Interact. Stud. **7**(3), 297–337 (2006)
2. M. Mori, The uncanny vall. Energy **7**(4), 33–35 (1970)
3. G. Hoffman, Ensemble: Fluency and Embodiment for Robots Acting with Humans, Ph.D. dissertation (2007)
4. A. Beck, L. Cañamero, K.A. Bard, Towards an affect space for robots to display emotional body language, in *The 19th IEEE International Symposium on Robot and Human Interactive Communication* (2010), pp. 464–469
5. H.G. Wallbott, Bodily expression of emotion. Eur. J. Social Psychol. **28**(6), 879–896 (1998)
6. T. Nomura, A. Nakao, Human evaluation of affective body motions expressed by a small-sized humanoid robot: comparison between elder people and university students, in *The 18th IEEE International Symposium on Robot and Human Interactive Communication* (2009), pp. 363–368
7. A. Kleinsmith, N. Bianchi-Berthouze, A. Steed, Automatic recognition of non-acted affective postures. IEEE Trans. Syst. Man Cybernet. Part B Cybernet. **41**(4), 1027–1038 (2011)
8. F.E. Pollick, H.M. Paterson, A. Bruderlin, A.J. Sanford, Perceiving Affect from Arm Movement. Cognition **82**, 51–61 (2001)
9. E. Crane, M. Gross, Motion capture and emotion: affect detection in whole body movement, in *Affective Computing and Intelligent Interaction* (Springer, Berlin, 2007), pp. 95–101
10. M. de Meijer, The contribution of general features of body movement to the attribution of emotions. J. Nonverbal Behav. **13**(4), 247–268 (1989)
11. J.M. Montepare, S.B. Goldstein, A. Clausen, The Identification of Emotions from Gait Information. J. Nonverbal Behav. **11**(1), 33–42 (1987)
12. C. Pelachaud, Studies on gesture expressivity for a virtual agent. Speech Commun. **51**(7), 630–639 (2009)
13. P. Gallaher, Individual differences in nonverbal behaviour: dimensions of style. J. Pers. Soc. Psychol. **63**(1), 133–145 (1992)
14. I. Poggi, Mind markers, in *Gestures, Meaning and Use*, ed. by N. Trigo, M. Rector, I. Poggi (University Fernando Pessoa Press, Oporto, Portugal, 2003)
15. C.L. Roether, L. Omlor, M.A. Giese, Features in the recognition of emotions from dynamic bodily expression, in *Dynamics of Visual Motion Processing: Neuronal, Behavioral, and Conputational Approaches*, ed. by U.J. Ilg, G.S. Masson (Springer, Boston, 2010), pp. 313–340

16. A. Beck, B. Stevens, , K.A. Bard, Comparing perception of affective body movements displayed by actors and animated characters, in *Proceedings of the Symposium on Mental States, Emotions, and their Embodiment* (2009)
17. K. Yamane, Y. Ariki, J. Hodgins, Animating non-humanoid characters with human motion data, in *Proceedings of the 2010 ACM SIGGRAPH/Eurographics Symposium on Computer Animation* (2010), pp. 169–178
18. A. Kleinsmith, N. Bianchi-berthouze, Recognizing affective dimensions from body posture, in *Affective Computing and Intelligent Interaction* (Springer, Berlin, 2007), pp. 48–58
19. G. Castellano, S.D. Villalba, A. Camurri, Recognising human emotions from body movement and gesture dynamics, in *Affective Computing and Intelligent Interaction* (Springer, Berlin, 2007), pp. 71–82
20. D. Bernhardt, P. Robinson, Detecting affect from non-stylised body motions, in *Affective Computing and Intelligent Interaction* (Springer, Berlin, 2007), pp. 59–70
21. K. Nakagawa, K. Shinozawa, H. Ishiguro, T. Akimoto, N. Hagita, Motion modification method to control affective nuances for robots, in *The 2009 IEEE/RSJ International Conference on Intelligent Robots and Systems* (2009), pp. 5003–5008
22. D. McColl, G. Nejat, Recognizing emotional body language displayed by a human-like social robot. Int. J. Social Robot. **6**, 261–280 (2014)
23. A. Yamaguchi, Y. Yano, S. Doki, S. Okuma, A study of emotional motion description by motion modification and adjectival expressions, in *The 2006 IEEE Conference on Cybernetics and Intelligent Systems*, vol. 4, pp. 1–6 (IEEE, New York, 2006)
24. J. Xu, J. Broekens, K. Hindriks, M.A. Neerincx, Mood expression through parameterized functional behavior of robots, in *The 22nd IEEE International Symposium on Robot and Human Interactive Communication*, vol. 22 (IEEE, New York, 2013), pp. 533–540
25. R.M. Murray, Z. Li, S.S. Sastry, in *A Mathematical Introduction to Robotic Manipulation* (CRC press, Boca Raton, 1994)
26. J. Broekens, W.-P. Brinkmand, AffectButton: a method for reliable and valid affective self-report. Int. J. Hum Comput Stud. **71**(6), 641–667 (2013)
27. M. Bradley, P.J. Lang, Measuring emotion: the self-assessment manikin and the semantic differential. J. Behav. Ther. Exp. Psychiatry **25**(1), 49–59 (1994)
28. M. Sawada, K. Suda, M. Ishii, Expression of emotions in dance: relation between arm movement characteristics and emotion. Percept. Mot. Skills **97**, 697–708 (2003)

Applications for Recording and Generating Human Body Motion with Labanotation

Worawat Choensawat, Minako Nakamura and Kozaburo Hachimura

Abstract Labanotation, invented for describing human body movement, is a dance notation system widely used in Western dance communities. The notation is very useful for composing a dance as well as performing it. This paper covers applications for recording and generating human body motion with Labanotation. The paper is divided into two parts: The first part describes a Labanotation editing tool named LabanEditor. LabanEditor includes the functionalities of both preparing Labanotation scores and displaying character animation so that beginners who are not familiar with Labanotation can study its description using a trial-and-error approach. The applicability of LabanEditor is shown by applying it to Noh-plays, a Japanese stylized traditional dance. The second part is an approach for generating Labanotation scores from motion capture data. The system consists of selecting key-frames of motion capture data, encoding posture for the key-frames, and generating Labanotation data. The experiments showed that for dancers, dance instructors and choreographers, both systems are potential tools for notating dance movements with Labanotation scores.

W. Choensawat (✉)
School of Science and Technology, Bangkok University, Multimedia Intelligent Technology Lab (BU-MIT), Bangkok, Thailand
e-mail: worawat.c@bu.ac.th

M. Nakamura
Graduate School of Humanities and Sciences, Ochanomizu University, Tokyo, Japan
e-mail: nakamura.minako@ocha.ac.jp

K. Hachimura
College of Information Science and Engineering, Ritsumeikan University, Kyoto, Japan
e-mail: hachimura@media.ritsumei.ac.jp

J.-P. Laumond and N. Abe (eds.), *Dance Notations and Robot Motion*,
Springer Tracts in Advanced Robotics 111,
DOI 10.1007/978-3-319-25739-6_19

1 Introduction

The dance community, mainly in Western countries, has widely accepted Labanotation as a graphical notation scheme for describing human body movement [14]. Dance notation, compared to motion capture, does not record the exact movements of a particular person, but it does record the essence of a movement so that anyone might interpret and perform this movement again. Moreover, symbolic representation has the advantage in that it enables us to record movement roughly and to make comparisons easily between various dances.

Recently, the digital recording and archiving of intangible cultural assets, such as classical dance and theatrical arts, has become an important research topic [11, 12]. Most composers of music prefer to record their compositions with music scores rather than playing them into an audio recording. The same is true for dance notation. In spite of the fact that Labanotation has many benefits, it also has several limitations as follows:

– Complexity: Labanotation is rich in symbols, and by using the full set of symbols almost all of body movements can be described. However, the resulting notation would become extremely complicated and difficult to comprehend.
– Time consuming: writing down the notation by observing the dance movement is a difficult and time consuming task which requires patience and skill.

The complexity issue is a question of the power of the description provided by Labanotation from a practical point of view. The problem is how to realize a method of describing detailed features and nuances of artistic, traditional dance movements while suppressing the complexity in the notation score.

This paper will cover an application named LabanEditor [6, 17] for preparing Labanotation scores and reproducing 3D CG character animation from the score using fundamental elements of Labanotation while maintaining the quality of the character animation.

For the solution of the time consuming issue, a system named GenLaban [7] for assisting users in generating Labanotation scores will be described. The automatic production of Labanotation scores can be achieved by using a motion capture system, which nowadays has been used for several purposes ranging from archiving the very precise data to real-time interaction. To utilize the motion capture data, GenLaban can generate Labanotation scores from it.

2 Labanotation

Labanotation is a graphical notation system for recording human body movements invented by Rudolf von Laban, an Austro-Hungarian dancer and choreographer, in the 1920s. A Labanotation score is drawn in the form of vertical staff where each column corresponds to a body part. The staff is read from bottom to top and the length of a symbol defines the duration of the movement. Similar to the music score, Labanotation uses bar lines to mark the measures as shown in Fig. 1a. The staff consists of three lines where the center line of the staff represents the center of the body: Columns on the right represent the right side of the body, and columns on the left, the left side of the body as shown in Fig. 1b. Figure 1b shows the basic arrangement of columns in the staff. The horizontal dimension of the staff represents the parts of the body, and the vertical dimension represents time.

Direction symbols are used to specify the direction of a movement. Symbols are placed in the columns of the staff. The basic direction symbols consist of nine horizontal symbols where the shape of a symbols represents the horizontal direction of movement of body parts. Shading within a direction symbol shows the level of a movement, i.e. vertical direction of movement (low, middle, and high), as shown in Fig. 2. Figure 3 shows right arm gesture according to the fundamental of direction symbols. From these fundamental principles, all movement can be recorded and additional symbols exist to record the full complexity of movement. Refer to [15] for more detail about Labanotation.

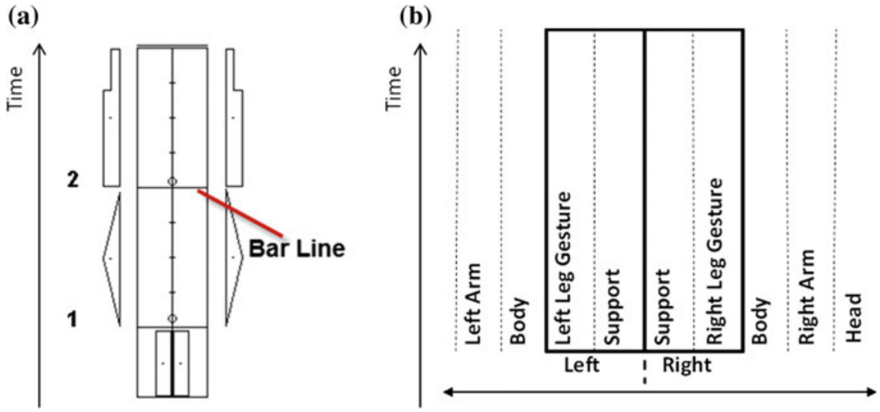
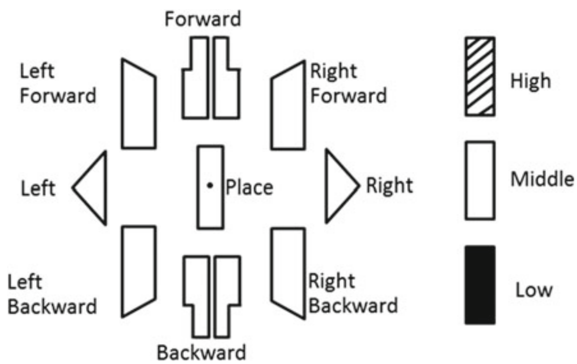


Fig. 1 Labanotation score. **a** Example of a Labanotation score and **b** columns of labanotation representing body parts

Fig. 2 Direction symbols; nine horizontal direction symbols and three kinds of shadings



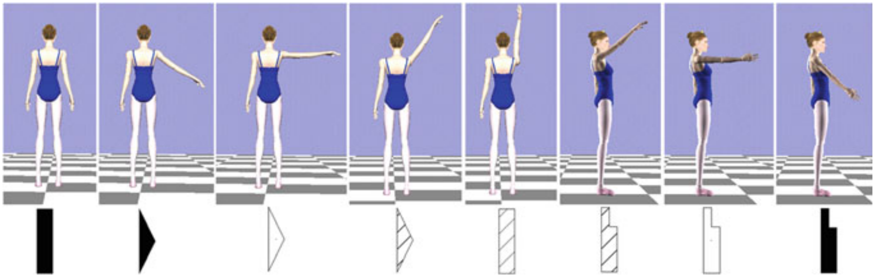


Fig. 3 Example of right arm gestures and its direction symbol

3 Systems for the Description of Human Body Movements

In teaching and studying the body motion of dance, documenting dancing can be accomplished in various media such as textual documentation, graphical notation, movies and videos. There are a number of studies on computerized dance notating systems.

The first work on the computer use of Labanotation was introduced by Brown and Smoliar in 1976 which initiated the design and implementation of a machine-independent data structure for Labanotation and its application in an interactive editing system [1].

LabanWriter [10] was developed at the Department of Dance at the Ohio State University. It is currently the most widely used Labanotation editor. The current version of LabanWriter is only for preparing Labanotation scores and recording them in digital form. It does not provide any functions for displaying character animations corresponding to the scores.

Kannan et al. [16] introduced a system named DanVideo which provides methods of annotation, authoring and retrieval tools for choreographers, dancers, and students. DanVideo annotations are defined by dance experts and produces MPEG-7 metadata. Unlike dance notation, the annotations in DanVideo describe the general properties of a movement but do not describe each specific detail of the movement such as body posture.

Soga et al. [19] developed a 3D animation system for the composition of ballet as a creation-support system for ballet teachers and students. The system allows users to create unique variations of enchainment by concatenating motion clips representing *pas* which are obtained by using motion capturing, but does not deal with Labanotation.

There have been several attempts to generate CG animation for representing dance Labanotation. The CG animation generator transforms Labanotation scores, which were prepared with LabanWriter, to the animation via the commercial software LifeForms [8]. However, LifeForms can only support the fundamental symbols of Labanotation.

LabanDancer [20] is a LabanWriter scores to 3D animation translation tool. Like LifeForms, LabanDancer does not have any functions for preparing Labanotation scores and supports only a limited number of symbols.

Since the above-mentioned Labanotation applications are separately designed and developed, there are no applications which can both create Labanotation scores and produce 3D CG character animation. For example, LabanWriter is able to input and edit the scores only, and LabanDancer is used only for displaying the movements.

4 LabanEditor

LabanEditor is an interactive graphical editor for editing Labanotation scores and displaying the 3D CG character animation associated with scores [6, 17]. LabanEditor integrates the following functionalities. Firstly, the 3D CG animation from a Labanotation score is shown directly after a user prepares a Labanotation score. Secondly, LabanEditor can import/export the data in LabanXML format, which is an XML representation of the Labanotation score.

LabanEditor consists of two major software components. The main component is for preparing the Labanotation score and the second, called MotionViewer is for displaying the 3D character animation. Both components allow users to input and edit the score and then display the 3D CG character animation immediately, which enables us to interactively confirm the movements.

Figure 4 shows the main window for editing the score and the MotionViewer for displaying the animation. While replaying the Labanotation score, users can observe the animation as well as the red horizontal line cursor, moving upward as the animation progresses, as shown in Fig. 4.

4.1 Conversion of Labanotation Scores to Computer Animation

In LabanEditor, Labanotation scores are represented in a simple format called Labanotation Data (LND), which uses alphanumeric characters to represent basic symbols. LND describes a pose of the body at each timing instance similar to key-frame body postures for animation, so that we can produce a motion of a body part by simply applying interpolation between the start and end key-frame poses. A key-frame pose of a body part is a moment of a pose at the end of a symbol. Figure 5a, b illustrate how a Labanotation score is converted to a LND structure. The lines that begin with "#" indicate the fundamental parameters of Labanotation. The movement of a body part is specified in the line followed by a command "direction", which corresponds to the Labanotation direction symbols.

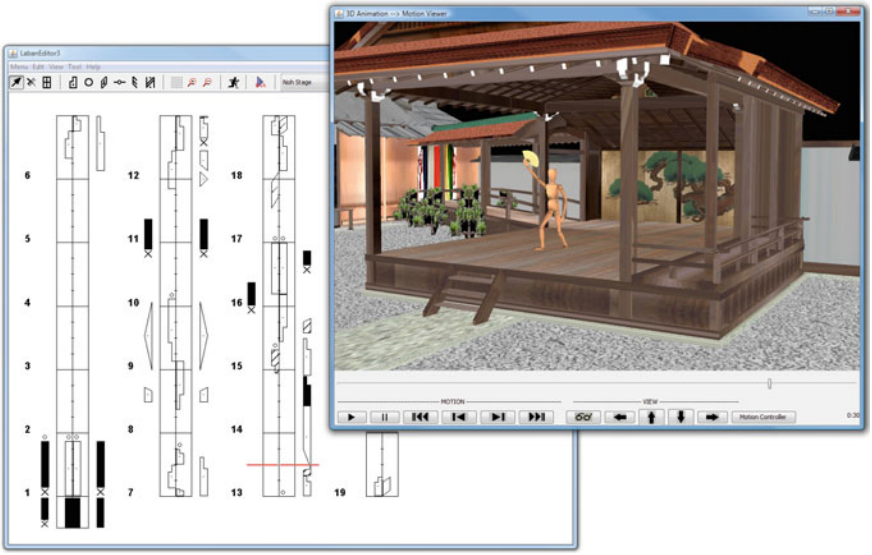


Fig. 4 LabanEditor user interface; The *left window* is the Labanotation editing window and the *right window* is the MotionViewer for displaying computer animation according to the score in the editing window

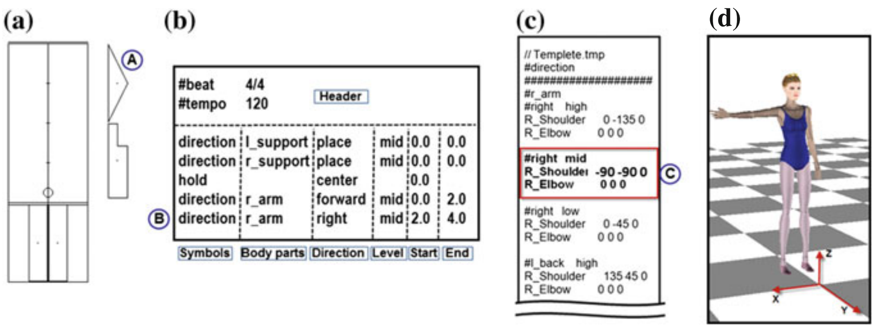


Fig. 5 Relationship between user input Labanotation symbols, LND and a template file; **a** user input score, **b** the LND representation of the score in (**a**), **c** part of a template file, and **d** the pose corresponding to the template in (**c**)

The system converts direction symbols into animation key-frames by using a template file which defines a mapping between the symbol and its corresponding pose of the body part. The template file describes the relationship between a direction symbol at the particular column and the rotation and translation of the corresponding joint. Figure 5 shows a notation, LND and description in a template file, and the resulting pose. The symbol marked Ⓐ in Fig. 5a is represented with LND as shown in the line marked Ⓑ in Fig. 5b. The LND line is mapped to the

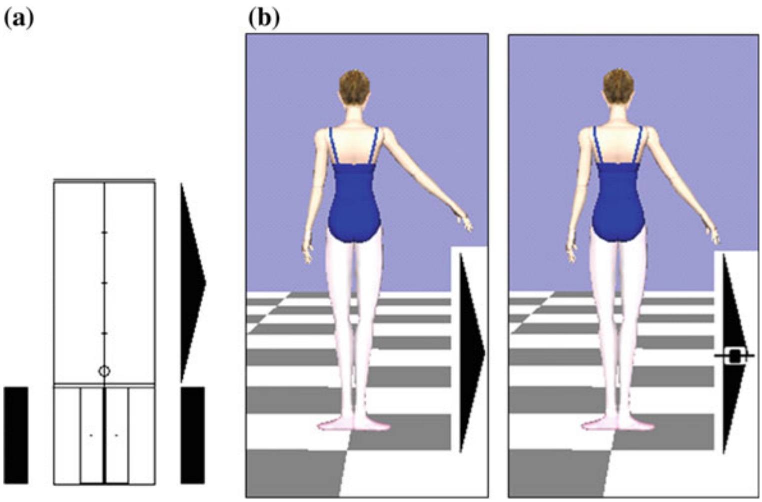


Fig. 6 Different poses expressed with a same direction symbol; **a** inputted scores and **b** two different poses corresponding to (**a**)

description of the part marked Ⓒ in the template file shown in Fig. 5c, which indicates a target pose of the right arm achieved by rotating the right shoulder joint 90° counterclockwise around the y-axis and 90° counterclockwise around the x-axis from the standard pose making the right palm facing forwards as shown in Fig. 5d.

When notating subtle movement of a dance, the score of Labanotation would be extremely complicated. For students new to notation, a complex score is a challenge to read and comprehend. However, when suppressing the complexity of Labanotation by using only fundamental Labanotation symbols, similar but distinct poses are sometimes defined with the same symbols. For example, the symbol in Fig. 6 illustrates that a performer moves his/her right arm to side low but at slightly different angles. Note that, these two poses can be differentiated by using more complex notation as shown in Fig. 6b, but for process of simplification, the 'root' symbol for both positions is side low, this then becomes the basic description or the fundamental symbol for representing these two different poses. A method of "dynamic templates" is used in our system in order to represent specific movements using a fundamental subset of Labanotation symbols only. These characteristic motions can be represented by changing the template files dynamically during the process of CG animation, while using very fundamental symbols.

The dynamic template method is to change templates dynamically depending on the motion to be described at the specified moment. For this purpose a template file manager is provided. Figure 7 shows an interface of the template file manager showing two template files are inserted and placed in the corresponding start time.

The information of the template files are inserted into an LND file corresponding to the start time as shown in Fig. 8b. The symbol "#include" determines the template file used at a particular timing. As a result, the Labanotation score shown

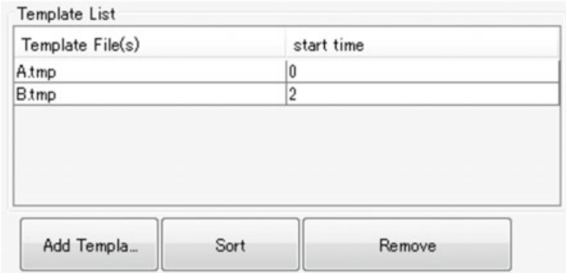
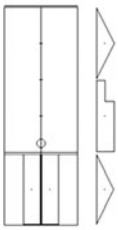


Fig. 7 Template file manager

(a)



(b)

```
#version  1
#beat     4/4
#tempo    120
#include        A.tmp
direction       l_support     place     mid     0.0  0.0
direction       r_support     place     mid     0.0  0.0
direction       r_arm         right     mid     0.0  0.0
hold            center                          0.0
direction       r_arm         forward   mid     0.0  2.0
#include        B.tmp
direction       r_arm         right     mid     2.0  4.0
```

Fig. 8 Example of the extension of LND file which supports the dynamic templates: **a** user input symbols and **b** the extension of LND file of the score in (**a**) supporting two motion templates

in Fig. 8a will be interpreted as the LND file shown in Fig. 8b. During the animation production process, the Labanotation symbols, represented in the format of LND, are mapped to the key-frame poses indicated by the current template.

4.2 Autonomous Dance Avatar

Body motion in ballet is usually composed of a sequence of many unit motions, *pas*, which are very characteristic to ballet. This will also apply to Japanese Noh play, in which the unit motions are called 'Kata' or form. It is said that there are more than 55 Katas in Noh play. If we can represent these Kata motions with simple Labanotation representation we can build a whole Noh play by concatenating these Katas on a time line. However, the representation in Labanotation would be very complex, when we want to represent the detailed characteristic body motions in Noh play with Labanotation.

As we mentioned, the Noh as well as Ballet are very stylized and we need not describe the detailed body motions. We can use one simple Labanotation representation without details as a representative of the stylized clip motions, Kata and *pas*. However, by doing so, a single Laban representation can have many possibilities of actual motions as shown in Fig. 6.

An actual body motion is represented as a sequence of poses, and the variation in in each pose from the standard pose, can be expressed by using a dynamic motion template mechanism. These dynamic templates can be prepared according to the characteristics of each dance genre. Dancers' personality can also be expressed with customized motion templates.

Then what we have to solve is to how to define a relationship between a simplified Labanotation clip and actual body motions represented with the sequence of poses.

Novice dancers would learn how to make a body motion of each particular unit motion clip being instructed by a dance teacher. We employed this analogy of teaching/learning method by using an associative memory system [2].

This section will explain a dance-style interpretation module embedded in the character model, called "Autonomous Dance Avatar" [5]. The embedded module enables the autonomous dance avatar to interpret the pattern of the Labanotation score and select an appropriate dance movement corresponding to the pattern.

In fact, the interpretation of dance styles is the recall process of stored Labanotation information and their associated motion templates. This can be achieved by teaching a character model to have a "dance-style memory", which is an associative memory storing relationship between Labanotation representations and the corresponding movements. The teaching process is called "training stage". Figure 9 illustrates the concept of the Dance Style Memory by using an Associative Memory System.

In the first training stage, some combinations of a Labanotation representation and actual body motion represented by a sequence of poses expressed with dynamic templates will be used as a training set.

In the recall stage when reproducing a dance motion from a Labanotation score, the associative memory is searched by entering a Labanotation representation of the unit motion. Because of the association mechanism of the associative memory we can get a proper motion clip which will suit the Labanotation input.

Considering the stored data, we will store a set of Labanotation symbols representing a pose and a sequence of poses representing a motion. The advantage of decomposing a Labanotation score into small units is to mimic the human learning system. Human beings can remember poses and a sequence of poses, and they still recognize poses with some differences.

In Labanotation, a posture (or pose) of the body can be represented by a combination of Labanotation symbols. A pose can be defined as a smallest unit of a movement. The problems for constructing the associative memory involve (1) how

Fig. 9 Dance style memory by using an associative memory system

to represent the graphical representation of Labanotation by some appropriate data formats and (2) how to store and recall motion templates in/from the associative memory.

4.2.1 Decomposition of a Labanotation Score into a Sequence of Poses

Motion, as recorded by Labanotation, can be broken down into a series of poses. This subsection will describe a method of breaking down a Labanotation score into a series of poses. Generally, the point of the end of each Labanotation symbol in a column will be a moment of a pose, but we have to think of the status of other columns at this moment. One endpoint of a symbol will contribute to make a pose. An example of describing poses is shown in Fig. 10, which the score contains four poses in a successive order, and each pose contains a set of symbols arranged at the same time. For a symbol, one symbol can belong to more than one pose. For example in Fig. 10, the right-arm-forward symbol resides in both poses $P2$ and $P3$. To decompose a Labanotation score into a sequence of poses, we employed a graph-theory problem.

The problem and its solution is described in the graph algorithms and applications as found in [4]. A graph is a representation of a set of objects where some pairs of objects are connected by links. The interconnected objects are represented by mathematical abstractions called vertices, and the links that connect some pairs of vertices are called edges.

Given a score, symbols appearing in the score can be represented by vertices. Let l_i, c_i, u_i be lower bound, middle point, and upper bound of symbol i, respectively. Figure 11 illustrates the definition of l_i, c_i, u_i. The edge between vertex i and j will appear if and only if the middle point of symbol i is in the upper and lower bound of the other symbol j. In the other words, edges e_{ij} will be one, if and only if

$$l_i \leq c_j \leq u_i \text{ OR } l_j \leq c_i \leq u_j. \tag{1}$$

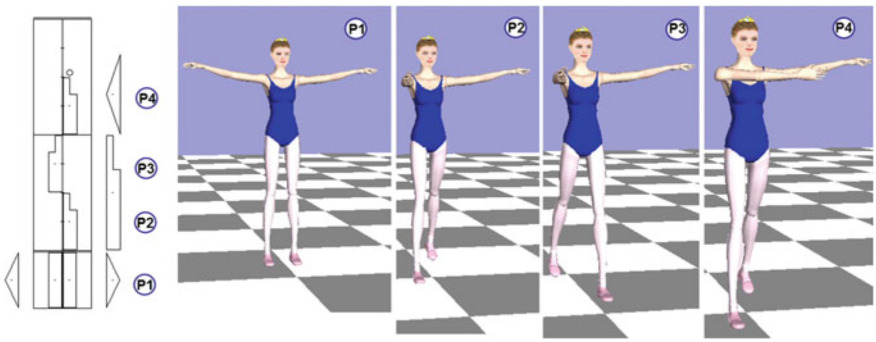


Fig. 10 Illustration of the poses in a score

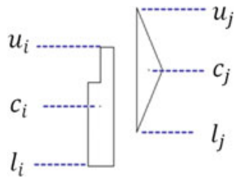


Fig. 11 Illustration of the definition of the lower bound, middle point, and upper bound of symbols i and j

Input : A Labanotation score
Output : Set of poses

1 *Draw an undirected graph G(V, E), where V is a set of all symbols appearing in the score;*
2 *Determine E = {e_{ij}} by using the equation below;*

$$e_{ij} = \begin{cases} 1 & if\ c_i \in [l_j, u_j] \\ 0 & otherwise \end{cases}$$

3 *Partition V into a minimum number of cliques;*

Fig. 12 Algorithm 1: Decomposing a Labanotation score to a set of minimum poses

When using the above description of transforming a score to a graph, a pose is represented as a maximal set of vertices where those vertices must be completely connected to each other. This is called *maximal complete subgraph* or *maximal clique*. Note that a clique is directly referred to a complete subgraph in term of Mathematics.

At this stage, we have successfully transformed our problem to the graph problem. The solution of our problem is to find a list of all maximal cliques, which means that we have to search the minimum number of cliques in a graph. The systematic approach will be written as shown in the Algorithm 1.

The Algorithm 1 in Fig. 12 shows how to break down a score into a number of poses by using an algorithm to find the minimum number of complete subgraph to cover all the graph's edges [3, 9], which a complete subgraph is a subgraph in which every pair of distinct vertices is connected by a unique edge.

For example, we will show a decomposition by a simple example. Given a Labanotation score as shown in Fig. 13, we can find four complete subgraphs i.e. poses by applying Algorithm 1 in Fig. 12. Table 1 shows the list of poses which are extracted from the Labanotation score in Fig. 13a.

4.2.2 Associative Memory for Autonomous Dance Avatar

The dance-style memory is implemented with a two-layered, associative memory. The first layer involves a recall process of known poses. Table 2 shows an example

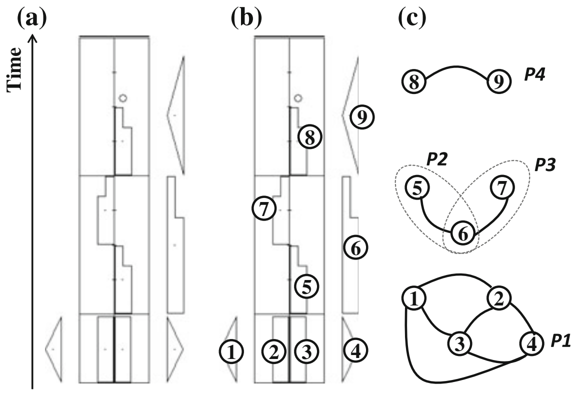


Fig. 13 Transform a Labanotation score to a graph. **a** Labanotation score. **b** Labeling symbols with 1 to *N* where *N* is a number of symbols in the score, and **c** the corresponding graph

Table 1 Example of a list of poses extracted from a Labanotation score

Labanotation score: Fig. 13a	
No. of symbols: 9	
No. of units/poses: 4	
Pose	Set of symbols
P_1	{1, 2, 3, 4}
P_2	{5, 6}
P_3	{6, 7}
P_4	{8, 9}

of a dance style interpretation with a combination of Labanotation symbols associated with pose#20 comprising three symbols of (l_support,place,mid), (r_support,place,mid), (r_arm,forward,mid). After that, the second layer classifies a sequence of poses to a trained motion template; for example, a sequence of pose#20 and #10 is classified as motion#3.

The approach consists of storing and recalling algorithms in Figs. 14 and 15, respectively. First of all, algorithms start with the decomposition of Labanotation scores into a sequence of poses, and then follow by training and recall stages as shown in Figs. 14 and 15, respectively. Algorithm 2 in Fig. 14 describes the implementation of an associative memory for storing poses and, then, motion templates which can be formed as a concatenation of poses while Algorithm 3 in Fig. 15 shows the recall method of motion templates.

In the recall stage, input data, a Labanotation score, can be different from the training data. Given an input Labanotation score, a set of motion templates can be assigned to the input score by applying Algorithm 2. After decomposing the score, Algorithm 2 starts with recalling a stored pose from the 1st-layer, associative memory for all poses as described in Fig. 15. For the 2nd-layer, associative memory, we adopt a concept of string matching for searching a set of motion templates since a sequence of poses is analogous to a text. That is to find an occurrence of a set of patterns in the text.

Table 2 Example of data stored in a two-layered associative memory

Layer	Query	Recall/recognition
1st	*Set of Labanotation symbols*	*Pose no.*
	(l_support,place,mid)	pose#20
	(r_support,place,mid)	
	(r_arm,forward,mid)	
	(l_arm,left,low)	pose#10
	(r_arm,forward,high)	
2nd	*Sequence of poses*	*Movement*
	(#20, #10)	motion#3
	(#5, #7, #30)	motion#8

Fig. 14 Algorithm 2: Developing the dance-style pattern storage

Input : A set of Labanotation scores and its corresponding motion templates
Output : A storage of dance-style patterns

1 *Decompose Labanotation scores into poses;*
2 *Create a two-layered, associative memory* for *storing dance-style patterns;*

1st : storing a set of Labanotation symbols and its associated poses
2nd : storing a sequence of poses and its associated motion templates

Fig. 15 Algorithm 3: Recall a motion template from the storage

Input : A Labanotation score
Output : A set of motion templates associated with its score

1 *Decompose Labanotation scores into poses;*
2 ***for*** *each pose* ***do***
3 | *Recall a most matching pose from the 1st-layer associative memory;*
4 ***end***
5 Recall a set of motion templates from the *2nd-layer associative memory;*

4.3 Use of LabanEditor for Noh Plays

This subsection demonstrates a case study of applying LabanEditor to traditional Japanese performing art, Noh. Noh plays are one of the most famous and characteristic Japanese traditional performing arts. Noh movements are highly stylized. 'Shimai' is a short but principal performance extracted from the whole Noh play. In principle, Shimai is composed of a number of prescribed movement units known as 'Kata', or form.

For the preparation of a dance style memory, the Noh plays were recorded using three video cameras in the following angles: front, side, and perspective views, respectively. A Shimai with 12 unique Kata was recorded. By precisely observing the videos, each Kata with Labanotation and their associated motion templates was described. These Labanotation scores and their associated motion templates were used as the training set to build the dance style memory.

The performance of the autonomous avatar is evaluated by two experiments. The first experiment is to test the recall of the motion templates in the dance style memory when some part of Kata were modified. The second experiment is to test whether the avatar can perform in natural Noh motion when a given Labanotation representation is not in the training set.

In the first experiment, the average percentages of correct recall are 87.88, 81.81, and 66.67 % when the modification degrees are 10, 20, and 30 %, respectively. The results show that the model can recall the correct motion template even thought the input score is not a prefect match to the training data.

In the second experiment, we got proper results where the autonomous dance avatar can perform an untrained Kata (not included in the training data) in natural Noh motion (Fig. 16).

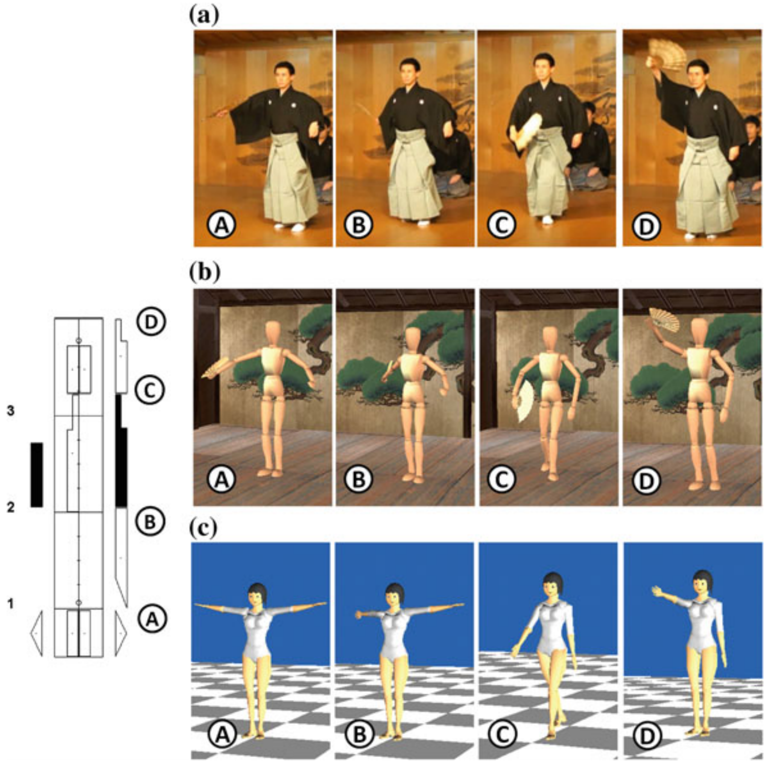


Fig. 16 Snapshots of the CG animation for a Labanotation score comparing between (**b**) the dance avatar with Noh knowledge and (**c**) the normal avatar. **a** Noh player performs 'Nisokude' Kata. **b** Dance avatar with Noh knowledge. **c** Dance avatar without Noh knowledge

5 GenLaban

The previous section covers the application for preparing a Labanotation score and displaying the animation of the score. When notating a dance, users have to manually input symbols on the staff which is not an easy task for a novice or student new to notation. This section will explain a system named GenLaban for assisting users in generating Labanotation scores. Figure 17 shows the system diagram of GenLaban which consists of the following processes:

1. Acquisition of motion data
2. Selection of key-frames
3. Analysis of body postures
   (a) Quantization body parts
   (b) Analysis of joints bending
   (c) Determination of weight support and jump
4. Generating Labanotation data (LND)

5.1 *Data Acquisition*

The most commonly used formats of motion capture data are C3D, TRC, ASF/AMC and BVH. The ASF/AMC and BVH formats store motion data in hierarchical skeleton data, while the C3D and TRC format store motion data in the 3D coordinates system. Currently, GenLaban only supports motion capture data in the BVH format.

The BVH is a format representing body motions on the bases of recording change of joint angles, and the body structure is represented by a skeleton hierarchical structure with a root at the pelvis. Figure 18 shows the human body model of BVH structure that is used in GenLaban.

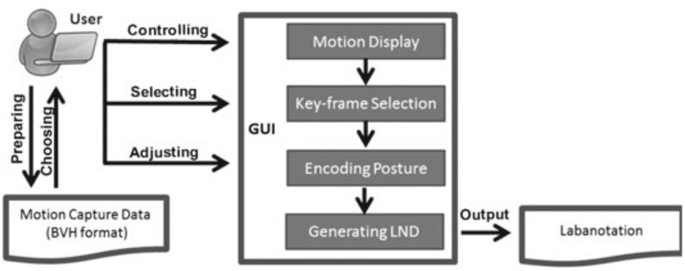


Fig. 17 System diagram of GenLaban

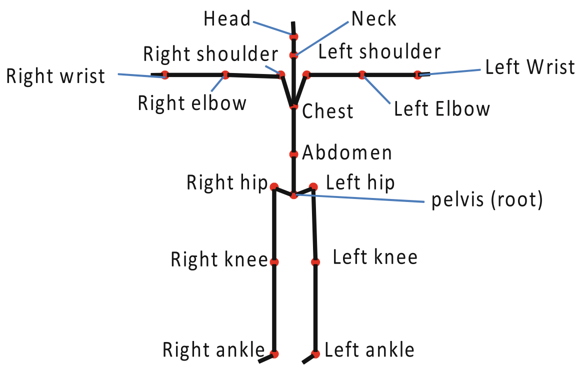


Fig. 18 Human body model in BVH format used in GenLaban

5.2 *Key-Frame Selection*

From Fig. 17, the generating LND consists of a motion capture file, together with some symbol definitions and suitable key-frame selection. Hachimura and Nakamura [13] introduced a method of key-frame selection using a threshold for the magnitude of the speed of the joint. However, for some types of dance especially a very slow one such as Noh-play, using a fixed threshold value for key-frame selection may not be appropriate since most of the time the magnitude of the joint speed will be below the threshold. Therefore, an interactive GUI for key-frame selection was developed. While displaying the motion capture file, the user can interactively select the suitable key-frames. Figure 19 shows a main user interface for key-frame selection where the motion capture file is displayed from three points of view as follows; (1) X–Y axis (front view), (2) Y–Z axis (side view), and

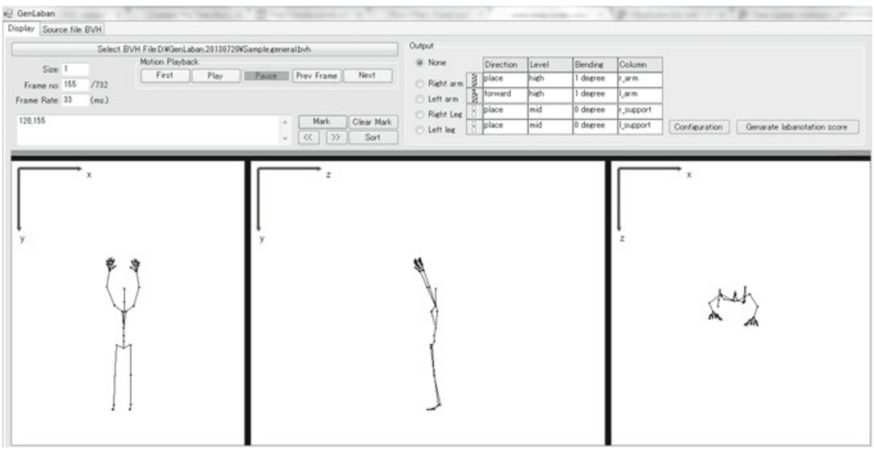


Fig. 19 Interactive graphical user interface of GenLaban

(3) X–Z axis (top view). The user can easily mark the key-frame and then generate the LND file from the marked key-frames.

5.3 Posture Encoding

To analyze the direction of movement of each body part, first we have to determine a "facing vector" which describes the direction of a performer's face. The facing vector is defined as the normal vector of a triangular plane defined through three points as follows: "right shoulder", "left shoulder" and "chest" as shown in Fig. 20a. Subsequently, the facing vector will be mapped into the standard cross of axes [15] which is the reference system of Labanotation most commonly used. In the standard cross of axes, the vertical axis, i.e. line of gravity, remains constant, and the forward direction in the horizontal axis is a modified facing vector that must lie at right angles to the vertical axis as shown in Fig. 20b.

5.3.1 Definition of a Motion Direction

After the forward vector is determined, the next step is to define the body posture of each selected key-frame in terms of Labanotation symbols. The quantization technique for the motion direction as proposed in [13] was adopted. The motion direction is defined as the direction from a parent joint to its child joint, which can be divided into 27 quantized spaces. The 27 spaces consist of 9 horizontal directions and 3 vertical levels as shown in Fig. 21.

For the motion direction of arms, a shoulder joint as a parent joint and wrist as the child joint is used. In a similar manner for a leg, the hip and its ankle joints are defined as the parent joint and the child joint, respectively.

We consider the horizontal distance between a parent joint and its child joint for *place*. If the distance is less than a predefined threshold, the child joint will be considered as *place*; otherwise, it must be classified into 8 directions as shown in Fig. 21a.

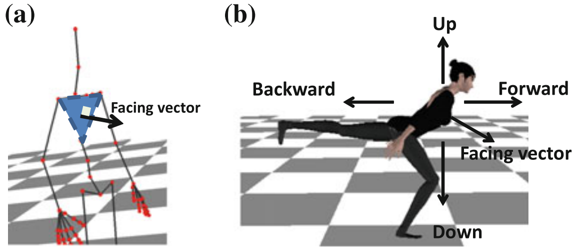


Fig. 20 Facing direction. **a** A normal vector defined as the facing direction, and **b** the standard cross of axes

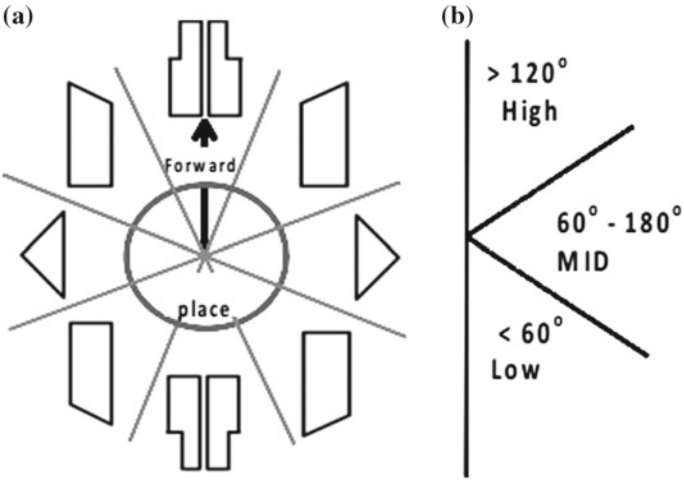


Fig. 21 Quantization of direction; **a** horizontal direction associated with forward direction, and **b** vertical direction according to the line of gravity in a standard cross of axes

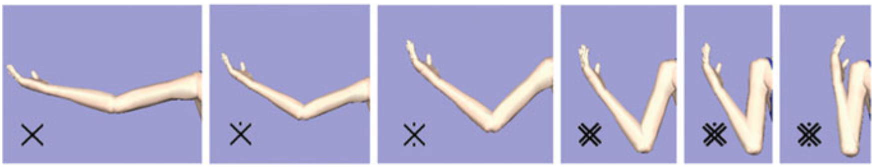


Fig. 22 Example of arm bending and its associated sign

5.3.2 Bending Analysis

In addition to the direction symbols, six signs are used to specify the bending of a body part as shown in Fig. 22. Considering the bending of the left elbow, the bending degree is calculated by considering three joints: the left shoulder(S), the left elbow(E), and the left wrist(W). A bending degree at elbow θ is derived by using two vectors $\overrightarrow{\mathbf{ES}}$ and $\overrightarrow{\mathbf{EW}}$, where $\overrightarrow{\mathbf{ES}}$ and $\overrightarrow{\mathbf{EW}}$ are the vector from an elbow to a shoulder, and the vector from an elbow to a wrist, respectively.

$$\theta = \arccos\left(\frac{\overrightarrow{\mathbf{ES}} \cdot \overrightarrow{\mathbf{EW}}}{\left|\overrightarrow{\mathbf{ES}}\right|\left|\overrightarrow{\mathbf{EW}}\right|}\right) \tag{2}$$

Given a fixed set of thresholds, a bending degree can be classified into six bending levels. The value of thresholds can be modified via the configuration window of the system.

5.3.3 Analysis of a Weight Support and Jump

When notating a dance, one of the most important processes is to determine which part of the body is carrying the weight. The body part carrying the weight must qualify for all of the following conditions:

1. The lowest body part must be at the ground level.
2. The lowest body part is stationary.

To determine whether a body part is stationary, the displacement of a particular body part between two adjacent key-frames is considered. In a case such as walking, for each walking step, one foot is always carrying the body weight; as a result, a foot can qualify with both conditions while the other foot may only qualify with one. Given a sliding walk as an example, the sliding walk is a walking style in which both feet can be at the ground level simultaneously. When we consider each step of the sliding walk, only the foot that supports the weight will remain stationary while the other foot slides forward.

When none of body parts follows with both of the conditions, it means no part of the body carries the body weight and the posture will be classified as a jump.

5.3.4 Determination of Duration

The next step is to determine the length of Labanotation symbols, which represent the duration of motion. The duration of a motion is calculated by the time between the previous and current key-frames. The formula for determining the symbol length in terms of beats is written as follows:

$$Symbol\,Length = \left(\frac{frame_c - frame_p}{frameRate}\right) \cdot \left(\frac{tempo}{60}\right) \tag{3}$$

where $frame_c$ and $frame_p$ are the frame number of the current and previous key-frames. A *frameRate* is frame rate of the motion capture data. The frame rate is expressed in frames per second (FPS). When reading a Labanotation score, *tempo* describes how fast or slow to perform a dance and is indicated in beats per minute (BPM).

5.4 Generating LND

By using the above method, every key-frame is transformed into successive LND lines, and consequently a series of many small Labanotation symbols is produced. The LND expressions determined so far are tentative because some may be redundant, i.e., the symbols for raising the left leg for half a minute may be repeated several times. Therefore, we can remove these redundant symbols by replacing them with the hold sign. The generated LND can be used to draw a Labanotation score and to display animation with LabanEditor.

5.5 Evaluation

In the first experiment, a Microsoft Kinect device was used as the motion capturing system. Kinect has a depth sensor that provides full-body 3D motion capture capability. Figure 23 shows the Kinect capturing full-body motion. The performer was asked to perform various types of arm and leg movements. Snapshots of the captured motion are shown in Fig. 24.

After capturing a motion, a set of key frames is manually selected, and the LND file is generated from those key-frames. Part of the generated LND data is shown in Fig. 25. The Labanotation score in Fig. 26 is then generated from the LND file using LabanEditor. Figure 27 shows snapshots of the CG animation of the Labanotation score in Fig. 26.

The second experiment is to illustrate the ability to handle motion when supporting weights are shifted. Walk and jump motions were selected from the motion capture data from CMU database [18].

For the example clip of walking, Fig. 28a shows the walking motion with the motion path represented by four trajectory lines for the right ankle, left ankle, right wrist, and left wrist joints. Figure 28b shows the resulting Labanotation score generated from the motion in Fig. 28a. The resulting score shows that GenLaban can determine the weight support of the walking motion correctly. For jump motion in Fig. 29, five key-frames are selected as shown Fig. 29a, b is the resulting Labanotation score generated from Fig. 29a. The resulting score shows that

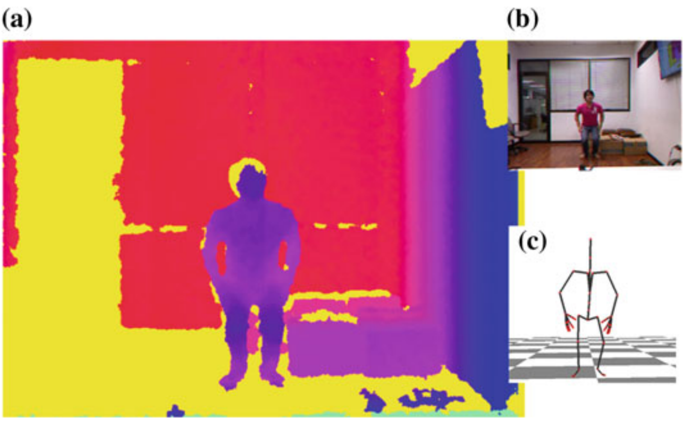


Fig. 23 Full-body 3D motion capture using Kinect; **a** the depth sensor, **b** motion capture, and **c** display of BVH model using the acquired 3-D information from (**b**)

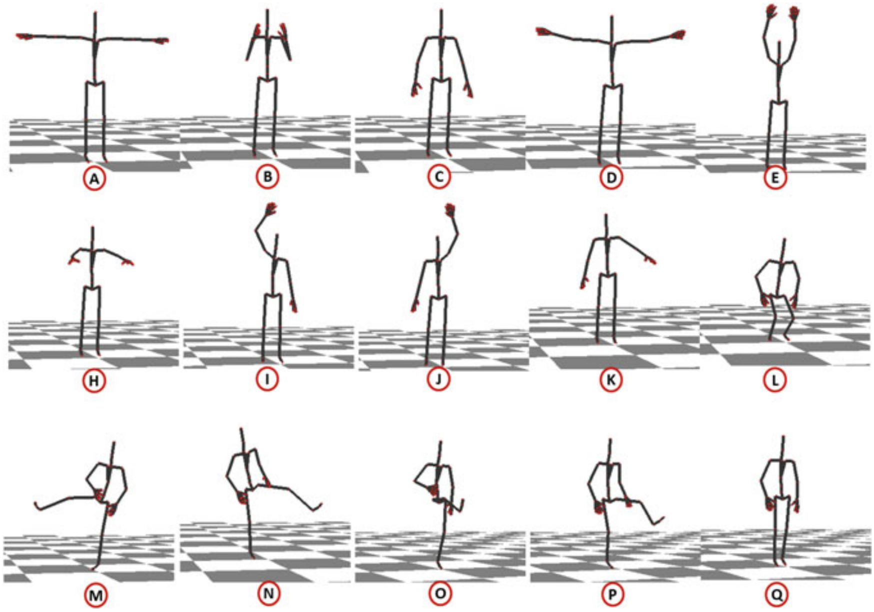


Fig. 24 The capturing motion

#beat	4/4						
#tempo	120						
direction	r_arm	right	mid	0.0	0.0		
direction	l_arm	left	mid	0.0	0.0		
direction	r_support	place	mid	0.0	0.0		
direction	l_support	place	mid	0.0	0.0		
direction	r_arm	place	mid	0.0	2.7	SPACE	-5
direction	l_arm	place	mid	0.0	2.7	SPACE	-6
direction	r_support	place	mid	0.0	2.7		
direction	l_support	place	mid	0.0	2.7		
hold	r_support			2.7			
hold	l_support			2.7			
direction	r_arm	forward	low	2.7	5.3	SPACE	-1
direction	l_arm	forward	low	2.7	5.3	SPACE	-1
direction	r_arm	right	mid	5.3	7.9	SPACE	-1
direction	l_arm	left	mid	5.3	7.9	SPACE	-1
direction	r_arm	forward	high	7.9	10.3	SPACE	-1

Fig. 25 LND file generated from motion in Fig. 24

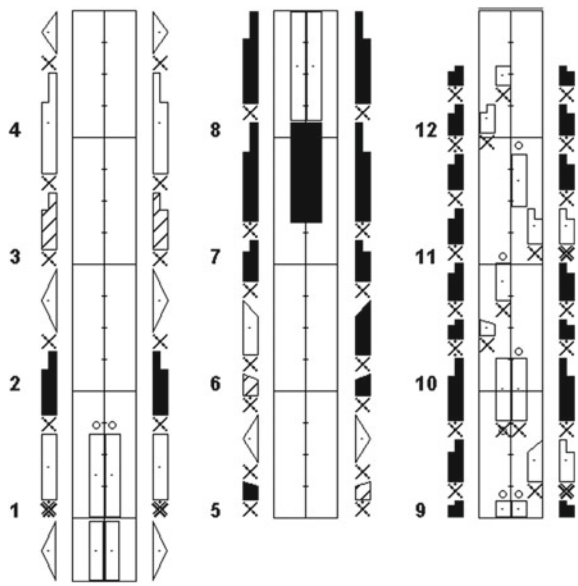


Fig. 26 Resulting Labanotation score

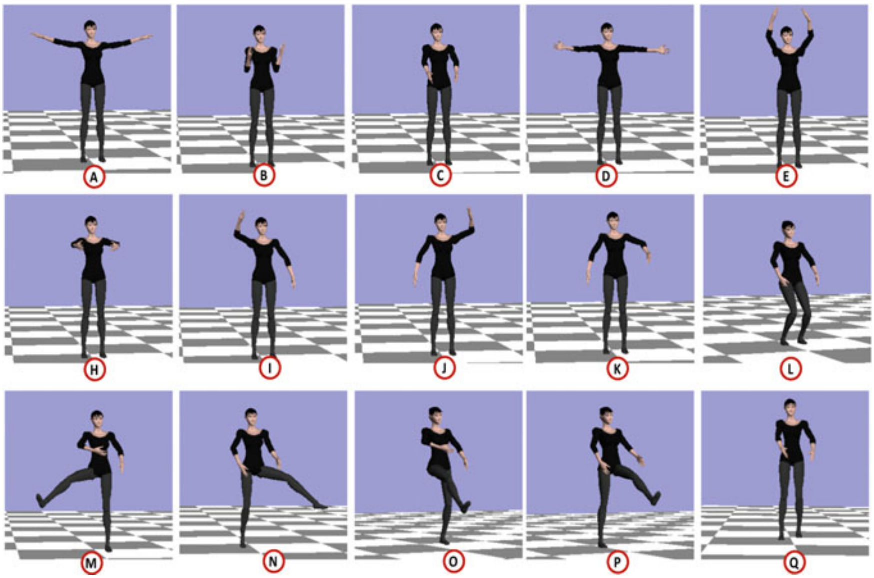


Fig. 27 Snapshots of 3D animation generated from the score in Fig. 26

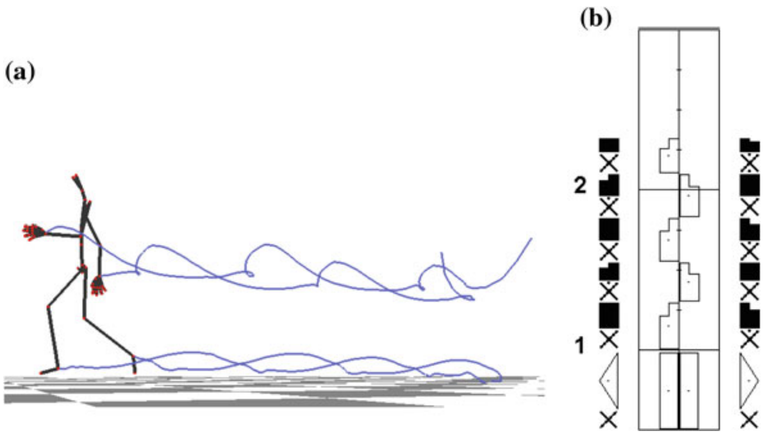


Fig. 28 The resulting Labanotation score from "walking" motion: **a** walking motion clip 08_01 where the motion path is represented by four trajectory lines and **b** Labanotation score associated with the motion clip in (**a**)

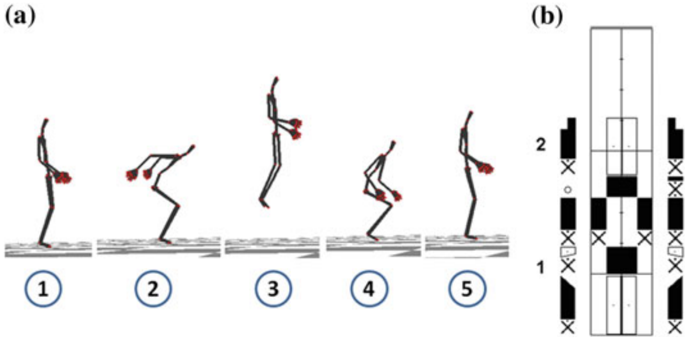


Fig. 29 The resulting Labanotation score from "jumping on the spot" motion: **a** five selected key-frames from motion clip 02_04 and **b** Labanotation score associated with the motion clip in (**a**)

GenLaban can determine the jumping motion correctly where no body part is carrying the weight.

This system has the ability to generate Labanotation score for choreographers who use this system to assist their work. However, in some cases, the generated scores are not grammatically correct; for example, the resulting score in Fig. 26 should have a hold sign after the final step. To deal with this problem, the users can use LabanEditor to make the correction of scores.

6 Conclusions

This paper has covered two applications for recording and generating human body motion with Labanotation. The first system is LabanEditor which is an interactive graphical editor for editing Labanotation scores and displaying the 3D CG character animation of the score. We presented the method for reproduction of stylized traditional dances such as Noh plays with the fundamental description of Labanotation called “Autonomous dance avatar”. The experiment showed that the Autonomous dance avatar can produce a natural Noh motion when given a Labanotation score. However, application of our method is not limited to Japanese dance, other stylized dances can also be handled.

LabanEditor will be beneficial to the following groups of people:

- Learners of Labanotation basics: after studying Labanotation basics with textbooks, learners can confirm the actual body motion by using character animation generated from the score.
- Dance researchers and dancers who want to use Labanotation in dance education/research: the system provides them a trial-and-error based learning method so that they can make their works more effective.
- Choreographers: the cycle of description of movement and successive confirmation of motion makes the choreographing process more effective.
- Masters who are teaching traditional dance such as Noh: they can explain and teach special body movements included in traditional performing arts like Noh.
- Novice dancers who want to experience traditional dance like Noh: they can continuously study body motions via the notation-animation cycle made possible by the system.

The second application is GenLaban for assisting choreographers to create the Labanotation score from motion capture data. The experiment demonstrates that the method and the quality of a Labanotation score are effective and usable. Because of the rough resolution of Labanotation, different users may make different judgments on symbol definition. The proposed tool provides a configuration panel for the user to define the symbol setting such as thresholds of the degrees of bending, directions of motion, and vertical levels.

GenLaban will be beneficial to the following groups:

- Novice dancers who begin to learn Labanotation: they can record their body motions and convert them to Labanotation.
- Dancers can record their body motions and convert them to Labanotation which can be used as a memory aid.
- Dance instructors who can use the system for preparing teaching materials.
- Choreographers can use the system for recording their ideas about the choreography of the performance.

The results show that the system is very useful for saving time to write Labanotation scores. However, the scores generated are not necessarily correct to all users. This is because the system allows a user to choose a set of key-frames by oneself; then, the generated scores will be different if a set of key-frames is not the same. Other reasons is that at present GenLaban can handle a subset of Labanotation which covers many of the fundamental movements. Therefore more complicated movements are require to be handled. However, Labanotation is rich in symbols and new symbols are being continually introduced. Supporting more symbols is basically a matter of time and attention to detail.

References

1. M.D. Brown, S.W. Smoliar, in *A graphics editor for labanotation*. ACM SIGGRAPH Computer Graphics, vol. 10 (ACM, 1976), pp. 60–65
2. A.G. Carpenter, Neural network models for pattern recognition and associative memory. Neural Netw. **20**(4), 243–257 (1989)
3. R. Carraghan, M.P. Pardalos, An exact algorithm for the maximum clique problem. Oper. Res. Lett. **90**(6), 375–382 (1990)
4. E. Cenek, L. Stewart, Maximum independent set and maximum clique algorithms for overlap graphs. Discrete Appl Math. **1310**(1), 77–91 (2003)
5. W. Choensawat, K. Hachimura, Autonomous dance avatar for generating stylized dance motion from simple dance notations. J. Inst. Image Electr. Eng. Jpn. **410**(4), 366–370 (2012)
6. W. Choensawat, S. Takahashi, M. Nakamura, W. Choi, K. Hachimura, Description and reproduction of stylized traditional dance body motion by using labanotation. Trans. Virtual Reality Soc. Jpn. **150**(3), 379–388 (2010)
7. W. Choensawat, M. Nakamura, K. Hachimura, Genlaban: a tool for generating labanotation from motion capture data. Multimedia Tools Appl. 1–24 (2014). doi:10.1007/s11042-014-2209-6
8. M. Coyle, S.D. Maranan, T. Calvert: a tool for translating dance notation to animation (2014), http://www.researchgate.net/publication/264121484_A_Tool_for_Translating_Dance_Notation_to_Animation. Accessed 22 July 2014
9. P. Erdös, Graph theory and probability. Canad. J. Math. **11**, 34G38 (1959)
10. I. Fox, R. Ryman, T. Calvert, Documenting dance for the 21st century: a translation interface between labanwriter and life forms, in *Twenty-second Biennial Conference of the International Council of Kinetography Laban*. July–2 August, vol. 26, pp. 193–205 (2001)
11. K. Hachimura, Digital archiving of dancing. Rev. Nat. Cent. Digitization (Online Journal) **8**, 51–66 (2006)
12. K. Hachimura, in *Digital Archiving of Dance by Using Motion-Capture Technology*, ed. by A. Kawashima et al. New Directions in Digital Humanities for Japanese Arts and Cultures, pp. 167–181 (2009)
13. K. Hachimura, M. Nakamura, Method of generating coded description of human body motion from motion-captured data, in *Proceedings 10th IEEE International Workshop on Robot and Human Interactive Communication, 2001*, pp. 122–127 (2001)
14. G.A. Hutchinson, *Labanotation* (Chapman & Hall, Routledge, 1977)
15. G.A. Hutchinson, *Labanotation: the system of analyzing and recording movement* (Routledge, 2013)
16. R. Kannan, F. Andres, C. Guetl, DanVideo: an mpeg-7 authoring and retrieval system for dance videos. Multimedia Tools Appl. **460**(2-3), 545–572 (2010)

17. K. Kojima, K. Hachimura, M. Nakamura, Labaneditor: Graphical editor for dance notation, in *Proceedings of 11th IEEE International Workshop on Robot and Human Interactive Communication, 2002* (IEEE, 2002), pp. 59–64
18. MoCap, CMU, The data used in this project was obtained from mocap. cs. cmu. edu. The database was created with funding from nsf eia-0196217 (2003)
19. A. Soga, B. Umino, T. Yasuda, S. Yokoi, Automatic composition and simulation system for ballet sequences. Visual Comp. **230**(5), 309–316 (2007)
20. L. Wilke, T. Calvert, R. Ryman, I. Fox, From dance notation to human animation: the labandancer project. Comp. Animation Virtual Worlds **160**(3–4), 201–211 (2005)

Human Motion Tracking by Robots

Katsu Yamane

Abstract Dance notation is a useful tool for providing a high-level description of complex movements to others. Given a description, human dancers are able to reproduce the intended detailed movements easily because they know how to move their joints to generate the described motion either through training or common sense. Robots, on the other hand, lack such training or common sense and therefore are incapable of creating fine details of the motion on their own. One possible way to "teach" a robot exactly how to move its joints is to use human motion data that include all the details. Unfortunately, using human motions for controlling robots is not straightforward due to various differences between human and robot bodies. This article will review some of the difficulties and give a brief survey on techniques for mapping human motions to robots.

1 Introduction

Humans can convert high-level descriptions of motions (words or other kind of symbols) into detailed joint movements. Unless the task is physically challenging (such as somersaults), this conversion is possible without difficulty because we have learned to perform those motions for years by observing others and practicing by ourselves. In other words, humans have a library of movements and controllers that can be adapted to create a motion matching the description.

Robots, on the other hand, are not equipped with such library, and therefore incapable of breaking down a high-level motion description into joint movements on their own. Instead, the developer has to "teach" them to acquire a library of natural and stylized motions.

One possible way to solve this problem is to provide a set of numerically optimized motion patterns or controllers. In this case, the library is given as a cost

K. Yamane (✉)
Disney Research, Pittsburgh, USA
e-mail: kyamane@disneyresearch.com

J.-P. Laumond and N. Abe (eds.), *Dance Notations and Robot Motion*,
Springer Tracts in Advanced Robotics 111,
DOI 10.1007/978-3-319-25739-6_20

function for optimization using the designer's intuition. In general, this approach works best for tasks whose objective can be easily formulated, such as minimizing energy consumption and maximizing speed. For example, walking [1] and running [2] have been generated through a simulated controller and optimization respectively. Dancing motions, however, would be difficult to generate with this approach because dance styles are often impossible to formulate.

Another way to solve this problem is to "teach" a motion from human performance, much like humans learn dancing by watching the instructor's movements. Motion capture data contain detailed information about human motions including body and joint trajectories. Unfortunately, it is not straightforward for a robot to imitate human motion due to various differences between the human and robot bodies. Human motion data are so specific to the performer's body that they usually require some adaptation to the robot's body. Similar differences also exist between a student and instructor bodies, which is why the student still needs practice to adapt the instructor's movement.

Figure 1 shows a conceptual representation of the discrepancy between notation (symbol) and physical movements, and the two approaches to fill in the gap. A notation (symbol) may include a wide variety of motions at the movement level, and only a small subset of them is desirable in terms of the style, context, etc. On the other hand, the robot is also able to perform only a subset of movements due to physical limitations. We have to find a motion that belongs to the intersection of these subsets.

Optimized motion or simulation can generate motions that are feasible for the robot, but may lack the desired style. Human motion data, on the other hand, would

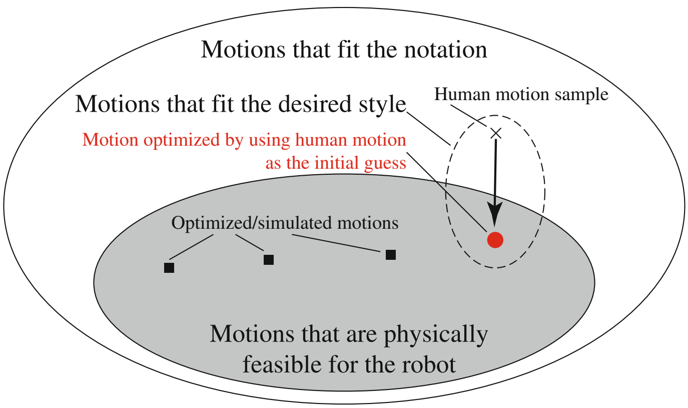


Fig. 1 The concepts used in this article. There are many possible motions that fit the notation (*largest oval*), but only some of them are feasible for the robot (*gray oval*) and even smaller subset meets the desired style (*oval with dashed line*). Motions generated by optimization or simulation would be feasible, but may not have acceptable styles (*black squares*). Starting from a human performance (*cross*), on the other hand, we may be able to obtain a motion that satisfies both feasibility and style constraints (*red circle*)

be esthetically desirable but may not be physically feasible for the robot. Using human motion data as the starting point for motion synthesis may have a higher probability of generating a motion that is both stylistic and feasible.

2 Differences Between Human and Robot Bodies

2.1 Measurement

First we will note some differences that arise from limitations in the current motion capture technologies. These topics are out of the scope of this article as we are interested in robot control not measurement systems, but we have to deal with these limitations nonetheless.

As with any measurement device, motion capture data suffer from noise and errors. In optical motion capture systems, for example, camera calibration error can cause error in the measured marker positions or, even worse, mislabeling of markers. Low-pass filters are often applied to overcome the noise issue but they also introduce delays and remove high-frequency movements. Another popular motion capture technology based on inertial (acceleration) measurement has the advantage of wide capture area, but they also suffer from the drift issue of inertial measurement units.

Most technologies require some form of markers attached to the human body, assuming that their movements accurately represent the human movement. However, those markers may move with respect to the bone, making the measured motion different from the actual skeleton motion. For example, even if markers are attached to the skin, the skin itself is not rigidly attached to the bone.

2.2 Body Kinematics, Dynamics, and Constraints

Physical robots' bodies are subject to physical constraints of those of available actuators and materials. Furthermore, every person has different body kinematics and inertial properties. It is therefore impossible to make a robot with exactly the same body kinematics and dynamics.

Some of the differences come from the inherent difference between biological and mechanical systems. For example, most of the human body elements are flexible while traditional mechanical systems are rigid. Actuators, whether electric, hydraulic, or pneumatic, also have very different properties from human muscles.

Other differences are due to physical constraints. It is often difficult to realize a hardware system with the same mobility as humans. For example, human spine has tens of degrees of freedom but most humanoid robots can only have three rotational joints due to space and weight restrictions. It is often difficult to realize the same range of motion of even normal humans.

For the same reason, it is impossible to realize the same mass distribution as humans. This difference would be problematic for motions that require precise balancing such as walking and running. It may even be impossible for a robot to perform the same motion as human performance without falling.

There are some robots that are designed to have human-like skeleton and actuator [3] but are still at experimental stage and difficult to produce at a large scale.

2.3 Kinematic Singularity

A non-intuitive issue with common robot design is the existence of kinematic singularity, often found at joints with multiple degrees of freedom (DOF). For example, the shoulder joint has three DOF: flexion/extension, abduction/adduction, and pronation/supination. In mechanical systems, three-DOF joints (spherical joints) are usually implemented with three rotational joints.

Figure 2 is a typical arm design that uses three joints for a shoulder joint. Consider a simple shoulder abduction motion where the arm is raised from the bottom while keeping the hand slightly in front of the body. Figure 3 illustrates the side and front view of the motion depicted at 22.5° interval of the shoulder abduction joint. The angles between the dashed lines in Fig. 3 represent the rotation of the flexion/extension joint. Notice that the flexion/extension joint has to turn rapidly when the arm is close to horizontal even though the abduction joint's speed is constant.

This phenomenon is caused by the kinematic singularity due to this specific shoulder joint structure. Even if the motion itself is easy to achieve for a spherical shoulder joint, the flexion/extension joint may exceed its velocity limit.

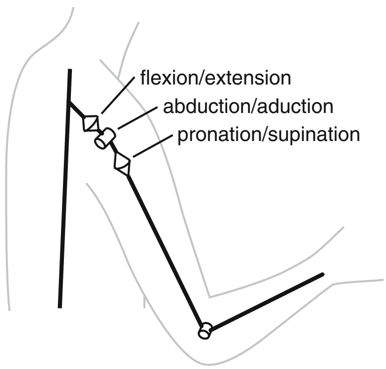


Fig. 2 An example of implementing a shoulder joint with three successive rotational joints

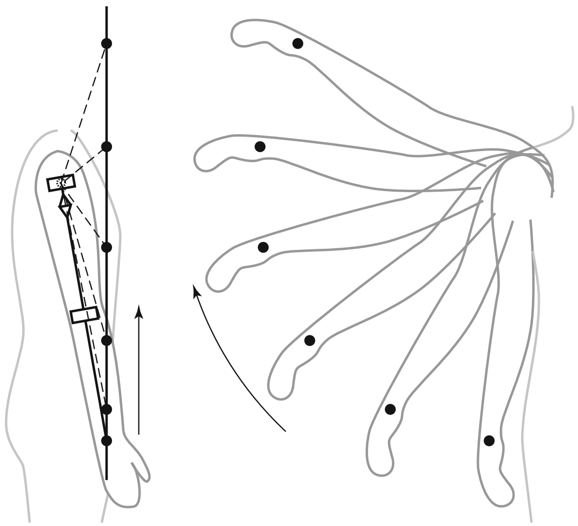


Fig. 3 Side (*left*) and front (*right*) view of the shoulder abduction motion that illustrates the singularity issue

3 Mapping Human Motions to Humanoid Robots

We will now review methods for mapping human motions to robots with human-like topology and proportions. Different methods focus on different aspects of the discrepancy between human and robot bodies.

3.1 Inverse Kinematics

In many motion capture techniques, we have to convert raw motion capture data to a robot skeleton motion represented as time series of joint angles. In optical motion capture, for example, the raw data are the three-dimensional trajectories of the markers. Converting 3D positions to joint angles is a well-studied problem in robotics called inverse kinematics (IK).

Traditional IK developed for robot manipulators considered the problem of converting an end-effector position to corresponding joint angles. Such methods do not work well for processing motion capture data because we have to consider 40–50 markers attached throughout the body. As a result, we have 120–150 dimensional measurement, which is far more than the DOF of typical humanoid robots. The data therefore may be inconsistent with humanoid skeleton due to error and noise as well as the additional DOF and flexibility of the human body. For such data, treating the marker positions as rigid constraints causes numerical issues.

A solution often employed is to treat the marker positions as soft constraints, i.e., minimize the total squared distance between the measured marker positions and corresponding markers attached to the robot's kinematics model. This solution can

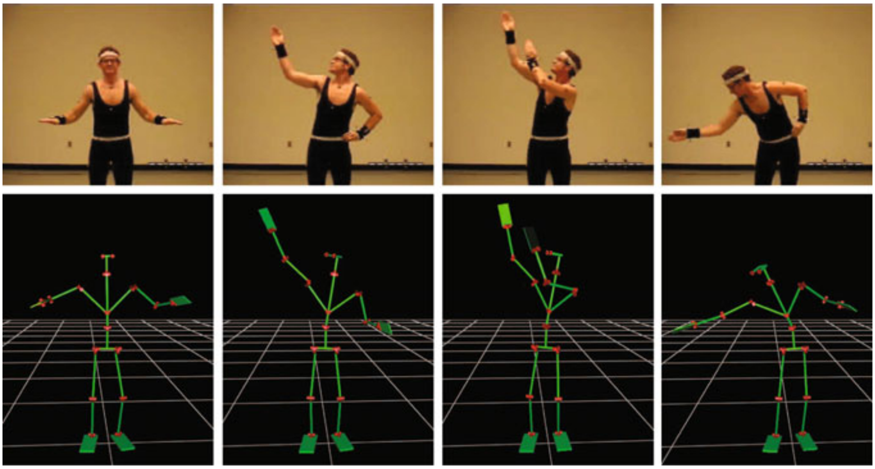


Fig. 4 Example of IK computation with marker position and joint motion range constraints

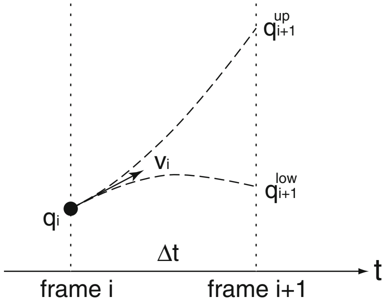


Fig. 5 Upper and lower bounds for the joint position at frame $i + 1$ computed from the previous joint position at frame i and the velocity and acceleration limits

also consider other constraints such as joint angle limits that can be considered as additional terms in the cost function.

Figure 4 shows an example of mapping human motion data to a robot's kinematic model. The most notable difference can been seen in the 2nd and last frames, where the robot's left hand does not touch the hip as in the human motion. This discrepancy is caused by the relatively small range of motion of the shoulder rotation joint.

Since this approach computes the robot pose at each frame separately, we cannot enforce velocity or acceleration limits. A workaround is to narrow down the joint position limits considering the previous joint position as well as the velocity and acceleration limits, as shown in Fig. 5. We may have to further shrink the range when the joint position is approaching the limit. Figure 6 the valid area in the position-velocity space to avoid exceeding the joint position limits in the future.

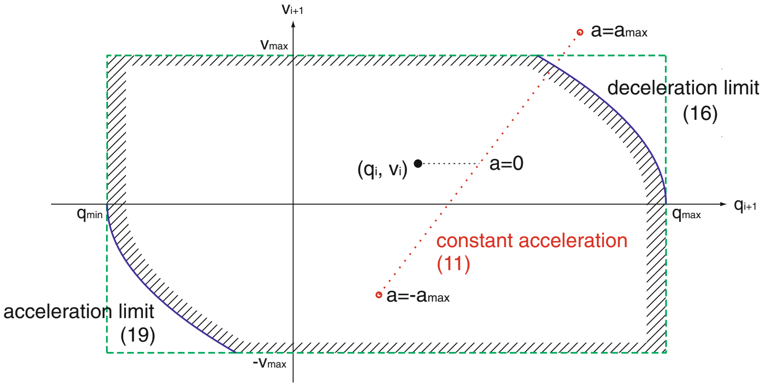


Fig. 6 An example of position-velocity limits to avoid future joint position limit violation, where point (p_i, v_i) is the current position and velocity and the *shadowed area* represents the valid area

3.2 Identifying Motion Feasibility

Once the motion is mapped onto the robot's kinematics model, we have a set of reference trajectories for the joints that can be tracked by a local feedback controller such as proportional-derivative (PD) controller. Unfortunately, simply tracking the reference trajectory at each joint does not always work due to one or more of the following reasons:

- The required joint torques exceed the actuator limits.
- The required contact force is infeasible.

The first limitation comes from actuator capability as well as robot weight. The second limitation applies to free-standing robots that use contact forces from the environment, such as in the case of walking. A contact force being infeasible means that the environment (typically the floor) has to pull the robot, or the friction force exceeds the limit determined by the normal force (friction cone; see Fig. 7).

The first issue can be identified by a classical robotics computation called inverse dynamics that estimates the required joint torques for given reference joint angles, velocities, and accelerations based on the inertial properties of the robot.

The second issue is more complicated. A classical method for identifying contact force feasibility is to compute the total center of pressure (COP), or equivalently, the Zero Moment Point (ZMP), and check if it is always inside the contact convex hull [4]. The COP/ZMP is the point of application of a single pure force equivalent to the total contact force and moment required to realize the motion. Because the normal force at all contact points must be in the repulsive direction, the COP must be inside the convex hull formed by those contact points. The ZMP has been used in many humanoid motion synthesis methods [5, 6] but it does not consider friction force limits. Also, COP/ZMP cannot be computed if the contacts are not coplanar.

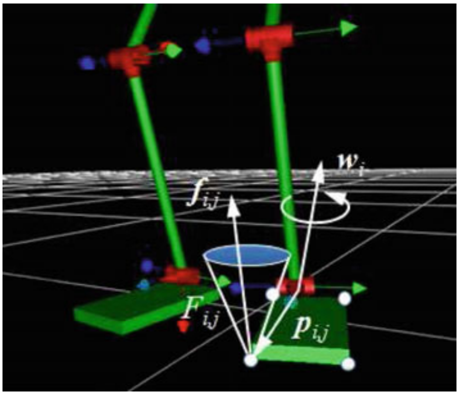


Fig. 7 Quantities related to contact forces; $f_{i,j}$: contact force, w_i: moment applied to the foot link by $f_{i,j}$ due to the position offset $p_{i,j}$, $F_{i,j}$: set of feasible contact force (friction cone)

Some researchers have proposed more general feasibility measure [7]. The author's group [8] proposed the general motion feasibility index that considers the both normal and friction force constraints, and used this index to generate feasible motions from captured data (see Sect. 3.3.1). This index is computed based on the distance from the required contact force to the "friction cone" surface in the six-dimensional wrench space. The index becomes negative when the required wrench is in the cone, i.e., feasible, and positive when in feasible.

3.3 Dealing with Infeasible Human Motion

If a motion is found to require infeasible contact forces, we need more sophisticated methods to make the motion feasible and prevent the robot from falling. This topic has been investigated in depth in graphics, as well as in robotics to a somewhat lesser extent. Most of existing methods can be divided into the following two categories:

- Modify the reference motion: if we modify the reference motion such that it is feasible for the robot, a local feedback controller is likely to be able to reproduce the motion. This controller is often combined with a simple balance controller so that the robot can cope with small disturbance.
- Use a balance controller: even if the reference motion is infeasible, we may be able to maintain the balance by using a balance controller. This approach is generally more robust to external disturbances.

3.3.1 Modify the Reference Motion

The first approach is to modify the human motion so that the new motion becomes feasible for the robot. This approach often assumes that there is a position tracking controller that allows the joints to accurately follow the modified motion.

The most obvious choice is to modify the poses that the robot goes through during the motion. For example, Nakaoka et al. [9] divided captured dancing motion into multiple tasks and optimized each task to be feasible for the robot dynamics. Miura et al. [10] computed feasible center of mass trajectory based on the center of pressure trajectory determined from the footsteps in the motion data, and then computed the joint motions by inverse kinematics. Compared to robotics, the graphics field has seen richer literature on this topic for wider range of motions [11–13] although the feasibility requirement tends to be less strict for graphics applications. Furthermore, virtual characters do not suffer from physical limitations such as maximum joint torques.

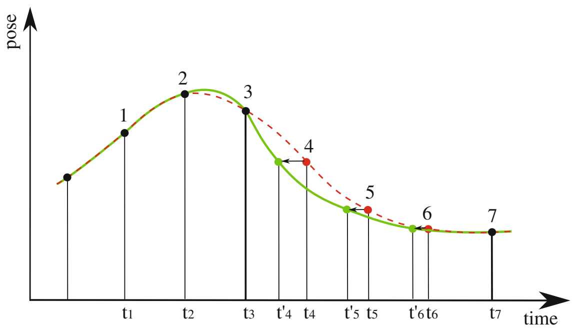
The author's group [8], on the other hand, developed a method for changing the timeline of the motion while keeping the same set of poses. This approach ensures that the robot passes through specific poses during the motion, which can be useful if the motion goes through a tight space and small change in the poses may case collisions.

Figure 8 illustrates the concept of time warping. Assume that we would like to keep the timing of frames 3 and 7 (because, for example, they are on the beat) but those of frames 4–6 can be changed. Changing the times of frames between 3 and 7 results in new velocity and acceleration profiles, which in turn changes the required contact forces and joint torques.

Figure 9 shows the motion feasibility index mentioned in the previous section at each frame of a stair climbing motion, where the dashed line denotes the original and solid the modified. The index must be negative at every frame for the motion to be feasible. As seen in the graph, some frames in the original motion violates the contact force feasibility but the entire modified motion is feasible.

Figure 10 shows the original (dark figure) and modified (bright figure) motions to highlight the differences. Note that the poses may be different at a particular time because we have modified the timeline.

Fig. 8 The concept of time warping



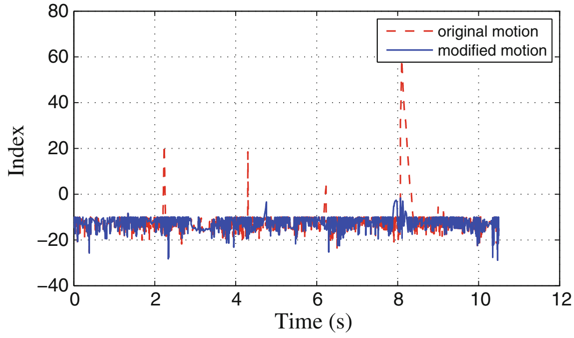


Fig. 9 General motion feasibility index of the original and modified motions of walking upstairs

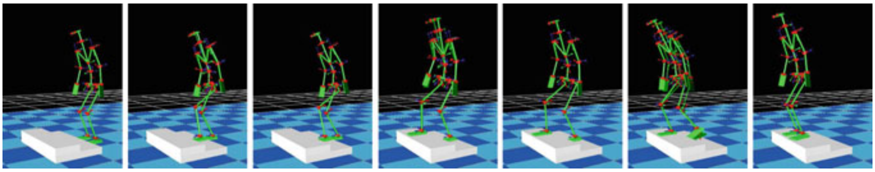


Fig. 10 Walking upstairs: original (*dark figure*) and modified (*bright figure*) motions

3.3.2 Combine with Balance Controller

Another approach is to use the data as is, but combine tracking with a balancing controller so that the robot does not fall down while tracking the human motion. For example, Ott et al. [14] focused on tracking upper body motions while maintaining the balance using the lower body. Again, a wider variety of controllers have been developed based on motion capture data in the graphics field [15, 16]. More recently, however, researchers appear to be focused on synthesizing human-like motions through optimization [17–19].

The author's group have developed a few techniques for tracking potentially infeasible human motions [20–22] used the whole body for both tracking and balancing. Figure 11 shows the overview of the controller first presented in [21]. In addition to the reference trajectory, the tracking controller takes the desired input (COP in this case) from the balance controller to maintain the balance. The tracking controller computes the joint torques that respect both tracking and balancing tasks.

Of course, we can also combine both approaches, namely use a balance controller along with a feasible motion to make the control easier. Figure 12 shows a simulation result of the balance controller in [22] applied to the time-warped feasible motion in Fig. 10. Note that the controller was not able to realize the climbing motion without time warping, and that the simulated motion is still slightly different from the time-warped, feasible motion.

Figures 13 and 14 show another example of walking downstairs.

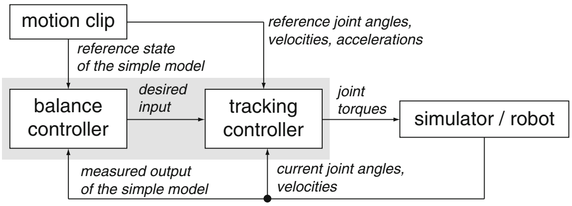


Fig. 11 Overview of the balance tracking controller

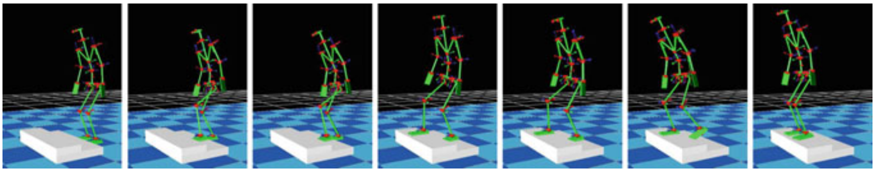


Fig. 12 Walking upstairs: time-warped (*dark figure*) and simulated (*bright figure*) motions

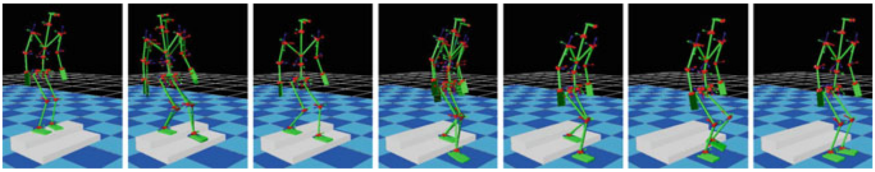


Fig. 13 Walking downstairs: original (*dark figure*) and modified (*bright figure*) motions

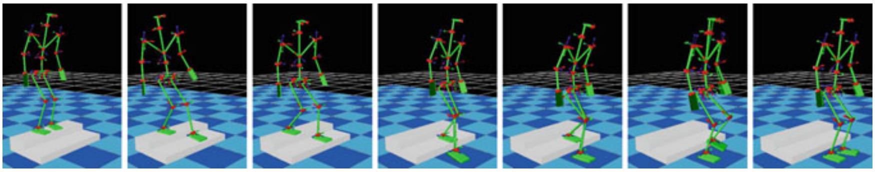


Fig. 14 Walking downstairs: time-warped (*dark figure*) and simulated (*bright figure*) motions

4 Model Versus Hardware

The final hurdle towards human motion tracking with robot hardware is the difference between a robot model and actual hardware. It is virtually impossible to obtain an accurate physical model of a robot due to inertial and kinematic parameter errors as well as unmodeled dynamics such as joint backlash and link deformation. As a result, many control or optimization algorithms that rely on a robot model do

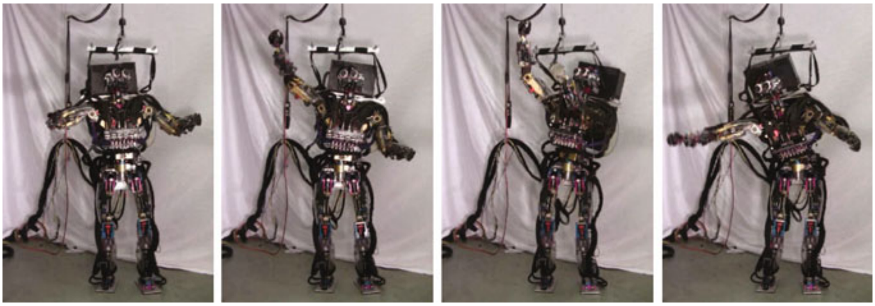


Fig. 15 Snapshots from the hardware experiment corresponding to the motion shown in Fig. 4

not work as expected from simulation and sometimes require control parameter re-tuning.

In general, tracking feasible trajectories with high-gain feedback control [9, 10] is less sensitive to model errors if sufficient margin from the contact convex hull boundary has been taken into account. Of course, they will also fail if the model error is so large that the motion becomes infeasible on the real robot. Controllers that use full-body dynamics model [21] tend to be more sensitive to model errors. It is also difficult to accurately estimate the inertial parameters of free-standing humanoid robots because we cannot obtain enough excitation necessary for accurate parameter estimation [23].

In the author's previous work [22], we corrected the dynamics model using actual joint torque and contact force measurements for hardware implementation of the controller [21], and realized the robot motion corresponding to Fig. 4 as shown in Fig. 15.

Another approach is to learn either the model or the controller through a number of experiments [24]. Some methods use human demonstration as the initial trajectory [25–27]. This approach is powerful because the robot can learn a controller that works for the actual hardware. However, conducting a large number of experiments (sometimes hundreds of them) is often difficult, especially for large-scale humanoid robots that have higher risk of breaking due to falls.

5 Summary and Future Directions

This chapter pointed out some of the issues related to the synthesis of fine details of robot motion from pre-recorded human motions, and reviewed several techniques that address these issues. In order to control a robot, symbolized notations have to be brought back to the continuous-time domain. Techniques similar to the ones introduced in this chapter can be used for controlling robots based on reference motions reconstructed from symbolized notations.

Tracking human motions with robot hardware, especially with humanoid robots, is not a solved problem. Many human motions are too fast to be tracked by current robot hardware to begin with, and robust humanoid balancing and locomotion control itself is an ongoing research issue. Furthermore, simply tracking pre-recorded human motion is not enough when the robot has to respond to the actual situation on the fly. However, balancing and locomotion tasks inherently require certain amount of look ahead (for example, you cannot lift a foot instantly; you have to shift your weight first) that makes some higher-level prediction necessary for realtime control. Another possible research direction is learning algorithms that can scale to complex humanoid robots but also work with small data set.

References

1. S. Coros, A. Karpathy, B. Jones, L. Reveret, M. van de Panne, Locomotion skills for simulated quadrupeds. ACM Trans. Graph. **30**(4), Article 59 (2011)
2. M. Posa, C. Cantu, R. Tedrake, A direct method for trajectory optimization of rigid bodies. Int. J. Robot. Res. **33**(1), 69–81 (2014)
3. Y. Nakanishi, Y. Asano, T. Kozuki, H. Mizoguchi, Y. Motegi, M. Osada, T. Shirai, J. Urata, K. Okada, M Inaba, Design concept of detail musculoskeletal humanoid "Kenshiro"—toward a real human body musculoskeletal simulator, in *Proceedings of the 2012 IEEE-RAS International Conference on Humanoid Robots* (2012), pp. 1–6
4. M. Vukobratovic, J. Stepanenko, On the stability of anthropomorphic systems. Math. Biosci. **15**, 1–37 (1972)
5. S. Kajita, F. Kanehiro, K. Kaneko, K. Fujiwara, K. Harada, K. Yokoi, H. Hirukawa, Biped walking pattern generation by using preview control of zero-moment point, in *Proceedings of IEEE International Conference on Robotics and Automation* (2003), pp. 1620–1626
6. T. Sugihara, Simulated regulator to synthesize ZMP manipulation and foot location for autonomous control of biped robots, in *Proceedings of IEEE International Conference on Robotics and Automation* (2008), pp. 1264–1269
7. H. Hirukawa, S. Hattori, K. Harada, S. Kajita, K. Kaneko, K. Fujiwara, M. Morisawa, A universal stability criterion of the foot contact of legged robots-adios ZMP, in *Proceedings of IEEE International Conference on Robotics and Automation* (2006), pp. 1976–1983
8. Y. Zheng, K. Yamane, Adapting human motions to humanoid robots through time warping based on a general motion feasibility index, in *Proceedings of IEEE International Conference on Robotics and Automation* (2015, in press)
9. S. Nakaoka, A. Nakazawa, F. Kanehiro, K. Kaneko, M. Morisawa, H. Hirukawa, K. Ikeuchi, Learning from observation paradigm: leg task models for enabling a biped humanoid robot to imitate human dances. Int. J. Robot. Res. **26**(8), 829–844 (2010)
10. K. Miura, M. Morisawa, F. Kanehiro, S. Kajia, K. Kaneko, K. Yokoi, Human-like walking with toe supporting for humanoids, in *IEEE/RSJ International Conference on Intelligent Robots and Systems* (2011), pp. 4428–4435
11. V.B. Zordan, J.K. Hodgins. Motion Capture-Driven Simulations that Hit and React, in *Proceedings of ACM SIGGRAPH Symposium on Computer Animation*, San Antonio, TX, July 2002, pp. 89–96
12. A. Safonova, J.K. Hodgins, N.S. Pollard, Synthesizing physically realistic human motion in low-dimensional, behavior-specific spaces. ACM Trans. Graph. **23**(3), 514–521 (2004)
13. K.W. Sok, M.M. Kim, J.H. Lee, Simulating biped behaviors from human motion data. ACM Trans. Graph. **26**(3), 107 (2007)

14. Ch. Ott, D.H. Lee, Y. Nakamura, Motion capture based human motion recognition and imitation by direct marker control, in *Proceedings of IEEE-RAS International Conference on Humanoid Robots* (2008), pp. 399–405
15. M. Da Silva, Y. Abe, J. Popović, Interactive simulation of stylized human locomotion. ACM Trans. Graph. **27**(3), 82 (2008)
16. U. Muico, Y.J. Lee, J. Popović, Z. Popović, Contact-aware nonlinear control of dynamic characters. ACM Trans. Graph. **28**(3), 81 (2009)
17. K.K. Yin, K. Loken, M. van de Panne, SIMBICON: simple biped locomotion control. ACM Trans. Graph. **26**(3), 105 (2007)
18. S. Jain, Y. Ye, C.K. Liu, Optimization-based interactive motion synthesis. ACM Trans. Graph. **28**(1), 10 (2009)
19. S. Coros, P. Beaudoin, M. van de Panne, Generalized biped walking control. ACM Trans. Graph. **29**(4), Article 130 (2010)
20. K. Yamane, Y. Nakamura, Dynamics filter-concept and implementation of on-line motion generator for human figures. IEEE Trans. Robot. Autom. **19**(3), 421–432 (2003)
21. K. Yamane, J.K. Hodgins, Simultaneous tracking and balancing of humanoid robots for imitating human motion capture data, in *Proceedings of IEEE/RSJ International Conference on Intelligent Robot Systems*, Oct 2009, pp. 2510–2517
22. K. Yamane, S.O. Anderson, J.K. Hodgins, Controlling humanoid robots with human motion data: experimental validation, in *Proceedings of IEEE-RAS International Conference on Humanoid Robots*, Nashville, TN, Dec 2010, pp. 504–510
23. K. Yamane, Practical kinematic and dynamic calibration methods for force-controlled humanoid robots, in *Proceedings of IEEE-RAS International Conference on Humanoids Robots*, Bled, Slovenia, Oct 2011
24. J. Kober, J.A. Bagnell, J. Peters, Reinforcement learning in robotics: A survey. Int. J. Robot. Res. **32**(11), 1238–1274 (2013)
25. C.G. Atkeson, S. Schaal, Robot learning from demonstration, in *International Conference on Machine Learning* (1997), pp. 12–20
26. J. Nakanishi, J. Morimoto, G. Endo, G. Cheng, S. Schaal, M. Kawato, Learning from demonstration and adaptation of biped locomotion. Robot. Auton. Syst. **47**, 79–91 (2004)
27. J. Kober, A. Wilhelm, E. Oztop, J. Peters, Reinforcement learning to adjust parameterized motor primitives to new situations. Auton. Robots **33**, 361–379 (2012)