

Antti Pirhonen
Stephen Brewster (Eds.)

LNC5 5270

Haptic and Audio Interaction Design

Third International Workshop, HAID 2008
Jyväskylä, Finland, September 2008
Proceedings

 Springer

Commenced Publication in 1973

Founding and Former Series Editors:

Gerhard Goos, Juris Hartmanis, and Jan van Leeuwen

Editorial Board

David Hutchison

Lancaster University, UK

Takeo Kanade

Carnegie Mellon University, Pittsburgh, PA, USA

Josef Kittler

University of Surrey, Guildford, UK

Jon M. Kleinberg

Cornell University, Ithaca, NY, USA

Alfred Kobsa

University of California, Irvine, CA, USA

Friedemann Mattern

ETH Zurich, Switzerland

John C. Mitchell

Stanford University, CA, USA

Moni Naor

Weizmann Institute of Science, Rehovot, Israel

Oscar Nierstrasz

University of Bern, Switzerland

C. Pandu Rangan

Indian Institute of Technology, Madras, India

Bernhard Steffen

University of Dortmund, Germany

Madhu Sudan

Massachusetts Institute of Technology, MA, USA

Demetri Terzopoulos

University of California, Los Angeles, CA, USA

Doug Tygar

University of California, Berkeley, CA, USA

Gerhard Weikum

Max-Planck Institute of Computer Science, Saarbruecken, Germany

Antti Pirhonen Stephen Brewster (Eds.)

Haptic and Audio Interaction Design

Third International Workshop, HAID 2008
Jyväskylä, Finland, September 15-16, 2008
Proceedings

Volume Editors

Antti Pirhonen
University of Jyväskylä
Department of Computer Science and Information Systems
40014 Jyväskylä, Finland
E-mail: pianta@jyu.fi

Stephen Brewster
University of Glasgow
Department of Computing Science
Glasgow G12 8QQ, UK
E-mail: stephen@dcs.gla.ac.uk

Library of Congress Control Number: 2008935109

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

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

ISSN 0302-9743
ISBN-10 3-540-87882-3 Springer Berlin Heidelberg New York
ISBN-13 978-3-540-87882-7 Springer Berlin Heidelberg New York

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

Springer is a part of Springer Science+Business Media

springer.com

© Springer-Verlag Berlin Heidelberg 2008
Printed in Germany

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India
Printed on acid-free paper SPIN: 12525805 06/3180 5 4 3 2 1 0

Preface

Bringing Them Under the Same Roof

The Haptic and Audio Interaction Design workshop series is now in its third year. These workshops have already demonstrated a clear need for a venue in which researchers and practitioners in these areas gather together under the same roof. Three years have also shown clear developments in the approaches taken – with the benefits of combining haptics and audio shown practically and conceptually in this year’s papers. In other words, it seems that when there is interaction between audio and haptic researchers, they really learn from each other and multimodal approaches emerge.

There are many good reasons for using haptics and audio together. There are the practical needs in application development. Mobile devices are an obvious example – while the device is small in size and is used on the move, interaction cannot rely solely on visual display. On the other hand, the development of applications for visually impaired people makes it necessary to learn how to design non-visual user-interfaces for different situations.

The rationale above may sound like haptics and audio are simply poor substitutes of visual-based interaction when vision is not an option. However, we argue that they have qualities which make them very different from visual interactions, making them more suitable for certain kinds of settings. The challenges are to identify where the benefits lie and then design interactions to take advantage of them. Perhaps the time is ripe to discard the naïve communication conceptions, in which we have the content and then just decide in which form or modality it is to be presented. The more we learn about the use of haptics and audio in human–computer interaction, the more effectively we can use them in design, whether visual displays are available or not. We believe that haptics and audio have the potential to change the entire way we interact with computational devices.

Thirteen papers were presented at the workshop, covering a wide range of research in the area of haptic and audio interaction. Alongside the papers we also had a poster and demonstration session, with 16 examples of haptic and audio design, to allow attendees to try out the new applications and interactions presented in the papers.

This year’s papers included very practical applications of the use of haptics and audio in interaction design. Design for visually impaired people was the focus of two papers. In the first (Pielot, Henze, Heuten and Boll), tactile belts were used to provide directional information. This is important as personal navigation is changing now that mobile phones commonly include GPS receivers. It is important to ensure that these new forms of navigation support are made available to all. The other paper in this category (Tanhua-Piironen, Pasto, Raisamo and Sallnäs) concerned group work applications, especially supporting the cooperation of sighted and visually impaired children in the school context.

The opportunities for audio-haptic integration were addressed in three papers, which considered very different kinds of applications. We learned how haptic feedback enhanced the use of expressive music controllers (Pedrosa and MacLean) and

how it could be combined with audio to help teach Tai-Chi (Portillo-Rodriguez, Sandoval-Gonzalez, Ruffaldi, Leonardi, Avizzano and Bergamasco). The third study of multimodal applications (Kryssanov, Kumokawa, Goncharenko and Ogawa) introduced a novel tool for navigating in social spaces, on the basis of information shared within a community.

An analytic approach to haptic and audio interaction is always challenging because the quantification of these modalities is inevitably complicated. An ambitious attempt to confront the challenge and mathematically model haptic interactions was presented by Hall, Rathod, Maiorca, Ioannou, Kazmierczak, O’Leary and Harris. In their paper, three different mathematical methods were empirically evaluated in terms of their appropriateness in classifying motion data. Another study which was based on empirical evaluation was highly qualitative in nature (Reis, de Sá and Carriço), investigating the effect of context of use on user preferences for interaction modalities with a mobile application.

What is then common between haptics and audio in interaction? Two different approaches for how haptics and audio could be conceptually handled together were presented. The first was based on esthetics frameworks (Chang and O’Sullivan), the other on the notion of physical embodiment (Pirhonen and Tuuri). This topic is important in providing foundations for designers to use to create successful multimodal interfaces.

Haptics and audio have a lot of potential in enhancing human–computer interaction, but new designs for interaction techniques that combine them effectively are needed if they are to be used. A tactile-sensitive surface element was first presented as a technical concept and then applied in sonification by Hermann and Kõiva. The second study in this category (Devallez, Rocchesso and Fontana) presented a tool which utilized depth cues in audio feedback connected to a gestural input device, thus providing a nice example of support to multimodal interaction.

Audio and haptics are both broad areas and there are still many basic perceptual questions to be answered about the two different modalities, in use on their own and in combination, before we fully understand them. New work was presented on the multimodal perception of rhythm (Jokiniemi, Raisamo, Lylykangas and Surakka) and roughness (Altinsoy), which will inform future interaction techniques and applications.

Having this collection of extremely different approaches to the discussion of haptic and audio interaction design was just a starting point for fruitful collaborations between researchers who participated in the workshop. We believe that the underlying idea of bringing together researchers from these different areas results in an interesting workshop and a creative exchange of ideas. The practical implementations of these ideas will be seen not only in scientific publications but, in the future, in everyday products as well.

September 2008

Antti Pirhonen
Stephen Brewster

Organization

The Third International Workshop on Haptic and Audio Interaction Design was organised by the University of Jyväskylä (Finland) and the University of Glasgow (UK).

Workshop Chairs

Antti Pirhonen	University of Jyväskylä (Finland), Department of Computer Science and Information Systems
Stephen Brewster	University of Glasgow (UK), Department of Computing Science

Poster and Demo Chairs

Andrew Crossan	University of Glasgow (UK), Department of Computing Science
Topi Kaaresoja	Nokia Ltd. (Finland)

Program Committee

Farshid Amirabdollahian	University of Salford, UK
Federico Barbagli	Stanford Robotics Lab/Hansen Medical, USA
Stephen Barrass	University of Canberra, Australia
Seungmoon Choi	POSTECH, Korea
Graeme Coleman	University of Dundee, UK
Abdulmotaleb El Saddik	University of Ottawa, Canada
Antonio Frisoli	PERCRO Laboratory, Scoula Superiore Sant'Anna, Italy
Stephen Furner	British Telecommunications Plc., UK
Matti Gröhn	CSC - Scientific Computing, Finland
Jing Hua	Wayne State University, USA
Gunnar Jansson	Uppsala University, Sweden
Johan Kildal	Nokia Ltd., Finland
Vuokko Lantz	Nokia Ltd., Finland
Charlotte Magnusson	Lund University, Sweden
Graham McAllister	University of Sussex, UK
Michael Miettinen	Suunto Ltd., Finland
Emma Murphy	McGill University, Canada
Manne-Sakari Mustonen	University of Jyväskylä, Finland
Ian Oakley	University of Madeira, Portugal

Sile O'Modhrain	Queen's University Belfast, UK
Antti Pirhonen	University of Jyväskylä, Finland
David Prytherch	Birmingham City University, UK
Roope Raisamo	University of Tampere, Finland
Chris Raymaekers	Hasselt University, Belgium
Sami Ronkainen	Nokia Ltd., Finland
Tony Stockman	Queen Mary University of London, UK
Kai Tuuri	University of Jyväskylä, Finland
Paul Vickers	Northumbria University, UK
Patrice L. (Tamar) Weiss	University of Haifa, Israel
Mark Wright	University of Edinburgh, UK
Wai Yu	Queen's University Belfast / Thales Air Defence Ltd., UK

The workshop was supported by two research projects:

- GEAR2, funded by Finnish Funding Agency for Technology and Innovation (www.tekes.fi) and the following partners: Nokia Ltd., GE Healthcare Finland Ltd., Sunit Ltd., Suunto Ltd., and Tampere city council.
- GAIME, funded by EPSRC (EP/F0230405), www.gaime-project.org.

Industrial Sponsors

The Nokia logo consists of the word "NOKIA" in a bold, blue, sans-serif font.The Suunto logo features a small red triangle above the word "SUUNTO" in a bold, black, sans-serif font.

Table of Contents

Visual Impairment

- Evaluation of Continuous Direction Encoding with Tactile Belts 1
Martin Pielot, Niels Henze, Wilko Heuten, and Susanne Boll
- Supporting Collaboration between Visually Impaired and Sighted
Children in a Multimodal Learning Environment 11
*Erika Tanhua-Piiroinen, Virpi Pasto, Roope Raisamo, and
Eva-Lotta Sallnäs*

Applications of Multimodality

- Perceptually Informed Roles for Haptic Feedback in Expressive Music
Controllers 21
Ricardo Pedrosa and Karon MacLean
- Real-Time Gesture Recognition, Evaluation and Feed-Forward
Correction of a Multimodal Tai-Chi Platform 30
*Otniel Portillo-Rodriguez, Oscar O. Sandoval-Gonzalez,
Emanuele Ruffaldi, Rosario Leonardi, Carlo Alberto Avizzano, and
Massimo Bergamasco*
- A System for Multimodal Exploration of Social Spaces 40
*Victor V. Kryssanov, Shizuka Kumokawa, Igor Goncharenko, and
Hitoshi Ogawa*

Evaluation

- Towards Haptic Performance Analysis Using K-Metrics 50
*Richard Hall, Hemang Rathod, Mauro Maiorca, Ioanna Ioannou,
Edmund Kazmierczak, Stephen O' Leary, and Peter Harris*
- Multimodal Interaction: Real Context Studies on Mobile Digital
Artefacts 60
Tiago Reis, Marco de Sá, and Luís Carriço

Conceptual Integration of Audio and Haptics

- An Audio-Haptic Aesthetic Framework Influenced by Visual Theory 70
Angela Chang and Conor O'Sullivan
- In Search for an Integrated Design Basis for Audio and Haptics 81
Antti Pirhonen and Kai Tuuri

Interaction Techniques

<i>tacTiles</i> for Ambient Intelligence and Interactive Sonification	91
<i>Thomas Hermann and Risto Kõiva</i>	
An Audio-Haptic Interface Concept Based on Depth Information	102
<i>Delphine Devallez, Davide Rocchesso, and Federico Fontana</i>	

Perception

Crossmodal Rhythm Perception	111
<i>Maria Jokiniemi, Roope Raisamo, Jani Lylykangas, and Veikko Surakka</i>	
The Effect of Auditory Cues on the Audiotactile Roughness Perception: Modulation Frequency and Sound Pressure Level	120
<i>M. Ercan Altinsoy</i>	

Author Index	131
-------------------------------	-----

Evaluation of Continuous Direction Encoding with Tactile Belts

Martin Pielot¹, Niels Henze¹, Wilko Heuten¹, and Susanne Boll²

¹OFFIS Institute for Information Technology, Germany

{pielot,henze,heuten}@offis.de

²University of Oldenburg, Germany

susanne.boll@uni-oldenburg.de

Abstract. Tactile displays consisting of tactors located around the user's waist are a proven means for displaying directions in the horizontal plane. These displays use the body location of tactors to express directions. In current implementations the number of directions that can be expressed is limited to the number of tactors. However, the required number of tactors might not be available or their configuration requires too much effort. This paper describes the design and the evaluation of a presentation method that allows displaying direction between tactors by interpolated their intensity. We compare this method with the prevalent one by letting participants determine directions and having them navigate along tactile waypoints in a virtual environment. The interpolated direction presentation significantly improved the accuracy of perceived directions. Discrete direction presentation, however, proved to be better suited for waypoint navigation and was found easier to process.

Keywords: multimodal user interfaces, tactile displays, direction presentation, interpolation, orientation and navigation.

1 Introduction

Maps and route descriptions are well established means for orienting in an unfamiliar area, keeping on track of a route, or finding points of interests (POIs). Most widely used tools for this are maps, either printed or digitally integrated in car navigation systems or mobile phones. These tools rely on the visual sense to be interpreted. In car navigation systems the visual display is complemented by a speech output which gives us directions to keep on track. However, both visual and auditory feedback might not be the most suitable ones to support a person's orientation and navigation while walking or driving. The visual display needs visual attention and is competing with the attention we need for watching and observing our surroundings. The auditory display can be perceived via speakers or earphones but can also be experienced as being annoying, obtrusive, and interfering with other tasks such as driving or talking to a friend.

In our research, we explored tactile sensation as a modality to present information for orientation and navigation. The driving argument behind this is that

tactile sensation can be perceived in a fashion that it is not obtrusive to the current user task and also can be perceived in discretion without the environment noticing it—as we know this from mobile phones that rest in vibration mode in our pockets. Tactile displays generally “appeal to the cutaneous senses by skin indentation, vibration, skin stretch and electrical stimulation.” [2]. Addressing the cutaneous sense, we developed and evaluated a tactile display that uses tactile transducers (tactors) worn around the waist. Their vibration can be sensed at different intensity levels and rhythms at the spot where they are attached to the skin. Using spatially distributed tactors each tactor’s location can be used as an additional output parameter. If spatially distributed tactors are used the tactile output at a specific location on the body can be connected to specific information.

Different groups [10,6,13,5] as well as ourselves [4] have shown that such tactile belts are a promising approach to provide directions in the horizontal plane. They used the tactors on the display to indicate a direction such as cardinal directions or the direction of waypoints. One drawback of the existing systems is, however, that the direction information is realized as a discrete presentation and each tactor conveys exactly one direction which results in either a high number of tactors or very coarse direction information.

In this paper, we present an evaluation of continuous direction presentation with a tactile belt. The continuous direction is encoded by intensity interpolation of adjacent tactors. Our evaluation bases on our tactile belt in which six tactors are distributed equally around a user’s waist. With two experiments we compare interpolated versus discrete tactile information presentation. The study provides evidence that interpolated direction presentation leads to a more accurate perception of the direction, while the discrete direction presentation was easier to perceive and process for waypoint navigation. The two major advantages of the interpolated direction presentation is that the accuracy of the perceived direction information increases. In addition, the system design can be parameterized such that existing tactile belts with varying numbers of tactors can now be re-programmed rather than re-engineered.

2 Related Work

Previous work has shown the feasibility of tactile displays for presenting directions by mapping them to body locations [1]. Many application scenarios have been suggested, like maintaining spatial awareness, waypoint navigation, and displaying the location of objects such as POIs.

Tan and Pentland [9] used a tactile belt for displaying cardinal directions to the user. This belt consisted of several tactors worn around the user’s waist and a compass. The system always activates the tactor that points most closely north. This kind of perception was evaluated by Nagel et al. [6]. For six weeks, four participants wore a belt which displayed north. Afterwards, a significant difference in a targeting task was observed between the experimental and the control group. Van Erp et al. [12] evaluated displaying directions for counteracting spatial disorientation. Participants were rotated around the yaw axis using a

swivel chair for 24 seconds. Afterwards, they had to compensate a quasi random angular velocity disturbance generated by the chair. Van Erp et al. showed that using a tactile display helps to recover spatial orientation.

Van Erp et al. [13] showed that pedestrians are able to follow a route consisting of waypoints guided by a tactile belt only. However, they observed that their participants were walking zigzag towards the waypoints. They argue that the limitation of factors resulted in a too inaccurate direction presentation. Tsukada and Yasumura [10] additionally found that reducing the length of vibration pulses has a negative effect on the perceptibility of directions. In our recent work [4] we could show that pedestrians can be guided close to an invisible route with a tactile belt when placing the waypoints very close to each other.

Lindeman et al. [5] proved the applicability of tactile belts for displaying stationary POIs. They evaluated user’s performance in a building clearing task, where a tactile belt helped them to avoid stepping into dangerous areas by displaying hazardous spots. In our previous work we proposed a system for keeping groups together in crowded environments [7]. A tactile belt is used to display the location of the group’s individuals. Rupert [8] investigated displaying pilots the location of objects around them in 3D space by using a tactile display consisting of 128 factors worn around the whole torso.

3 Direction Presentation with a Tactile Belt

In our previous work, we developed a belt type tactile display with six factors [4]. The belt consists of flexible fabrics and is worn around the hip. Six vibration motors serving as factors are sewn into the belt. The factors are composed of an unbalanced mass on a rotating axis and can produce vibrations of high frequencies. They are equally distributed leading to a distance of about 60° between two adjacent factors (see illustration in Figure 1).

The hardware design of our belt is basically comparable to the systems we presented in the related work section. In these systems, direction presentation using tactile belts typically follows a concept that we call ”discrete direction presentation”. Directions are expressed by modifying the body location of the tactile cue, where directions are mapped to body location. A factor is activated if the corresponding direction should be displayed to the user. As the number of factors is limited on the belt, a whole range of directions must be mapped to be displayed by one single factor. Hence, each factor is responsible for displaying a range of directions as illustrated in Figure 2.

This presentation method leads to an inherent inaccuracy. Taking our belt as example, each factor is mapped to a range of directions with a size of 60° . Thus, the deviation between actual and expressed directions lies between 0° - 30° . This results into an average deviation of 15° , assuming evenly distributed directions being displayed. One solution to this problem is to alter the mapping between directions and body location considering the application scenario. For waypoint navigation, one extra factor could indicate being on route, while the others are used to display the direction of the subsequent waypoint. However,

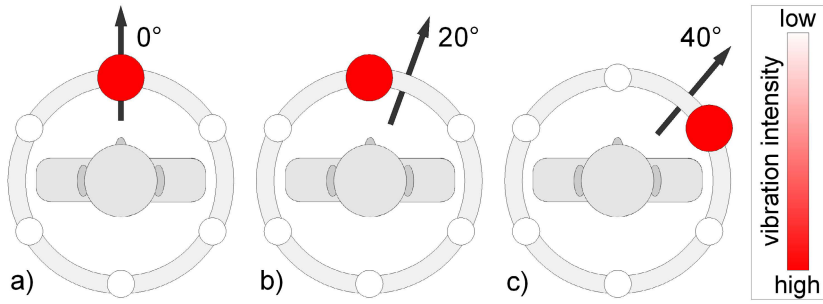


Fig. 1. Sketch of a belt with six tactors displaying three direction (0° , 20° , and 40°) using a discrete presentation technique

this is not suitable for other use-cases such as displaying the location of POIs that could be located all around the user. Therefore a solution would be adding tactors to gain higher accuracy. We propose an alternative approach for gaining higher accuracy which makes best use of number of available tactors by using the vibration intensity as additional parameter for encoding directions.

Combining the parameters body location and intensity we propose a presentation method that displays directions using interpolation. This encoding exploits the effect of *apparent location* where a single perceived stimulus is induced by two stimuli at different locations. According to van Erp [11] the perceived location depends on the relative magnitude of the two stimuli. Directions between the exact angles of two adjacent tactors are encoded through different intensity levels of these tactors. The intensity levels are determined by a linear function depending on the displayed angle. If two tactors are 60° apart, and an angle is displayed 20° away from one tactor (see Figure 2), the intensity of the closer tactor is two-third, and the intensity of the other tactor is one-third of the maximum intensity. We proposed this idea in [3] for a display with three tactors. In the following this method will be called interpolated presentation. This method allows making existing devices more accurate and flexible without the need for physical re-engineering. Consequently, this opens up new options for flexible software configuration of the belt's output.

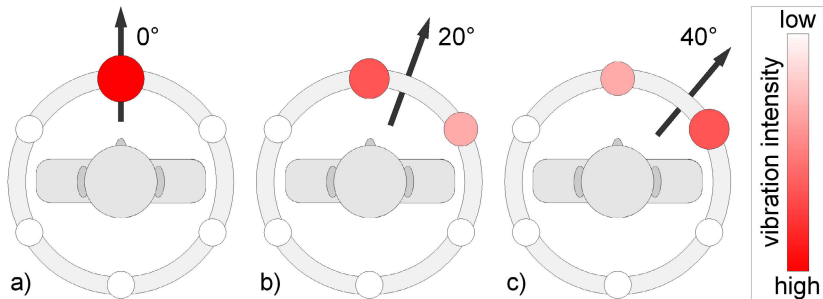


Fig. 2. Sketch of a belt with six tactors displaying three direction (0° , 20° , and 40°) using an interpolating presentation technique

4 Evaluation

We conducted two experiments to compare interpolated and discrete presentation. In the first experiment we asked participants to determine directions. In the second experiment participants had to navigate along tactile waypoints in a virtual environment. We expected people to perceive directions more accurate using the interpolated presentation. We also assumed that the interpolated presentation is less obtrusive, since when the displayed direction changes slowly, no sudden jumps of the tactile feedback from one tactor to the other occur. On the other hand, we supposed that discrete feedback is easier to interpret for untrained users, since there is less information to process. 16 participants with an age of 20-34 ($M^1=25.73$, $SE^2=1.05$) took part in the evaluation. 15 of them were male and eight had previous experience with our tactile belt. We afterwards handed out questionnaires to ask for the participants' subjective impressions. The experiments and the questionnaire are described in the following.

4.1 Accuracy of Direction Perception

The aim of the first experiment was to test the assumptions that interpolated presentation is more accurate but more difficult to process. We therefore let the participants determine a number of directions displayed with both presentation methods and compared the reaction time of the users as well as the accuracy of the perceived directions.

Method. Discrete presentation served as control condition and interpolated presentation as experimental condition. Every participant contributed to both groups. To rule out systematic sequence effects, we randomly assigned which display method was used first. We measured the average deviation of the determined directions for comparing the accuracy of the presentation methods. Additionally we recorded the reaction time as an indicator for the difficulty.

The tactile belt was connected to a desktop computer. An application running on the computer was used to display directions via the tactile belt. A circle on the screen enabled the participants to select the perceived direction using a mouse (see Figure 3). No other visual cues were given except a line marking front.

Prior to the experiments 32 random directions between 0° and 359° were generated. We displayed the same directions to every participant for comparability purposes. At the beginning of each experiment session, each participant was introduced to both display methods theoretically. Then, the application demonstrated both presentation methods by displaying a virtual point running clockwise around the participant. When the participants had familiarized themselves with the presentation methods, they were informed which method was used first to display directions. The application then presented the 32 directions in random order. Each direction was presented until the participant responded to the system by indicating the perceived direction. Afterwards, all directions were presented again in another random order using the other presentation method.

¹ Mean.

² Standard error.



Fig. 3. (a) Participant during the evaluation wearing the tactile belt. (b) Application to specify the perceived direction.

Results. The mean deviation of directions given by the participants was significantly lower with the interpolated presentation (16.83° , SE 0.74) compared to the discrete presentation (19.43° , SE 0.97). In contrast, the reaction time was significantly higher for the interpolated presentation (4.42s, SE 0.44) than for the discrete presentation (3.23s, SE 0.26). Statistically, interpolation had a medium effect on improving the accuracy ($t(14) = 2.93$, $p < .01$, $r = .49$) and a high effect on prolonging the reaction time ($t(14) = -4.54$, $p < .001$, $r = .84$). Comparing the performance of participants that had previous experience with the tactile belt to those who had none did not reveal any significant effect.

4.2 Waypoint Navigation in Virtual Environments

The goal of the second experiment was to test, how interpolated presentation performs in an exemplary task like waypoint navigation. We asked the participants to walk along a route in a virtual 3D environment. We measured and compared the time the participants needed to complete the route with each presentation method. We made no assumption about which presentation method would allow faster completion of the route. On the one hand, we expected that more accurate feedback would result in the ability to follow the ideal route more closely, but on the other hand, the expected higher difficulty of processing the feedback could nullify or invert this effect.

Method. The participants explored a virtual area from the first person view. They had to follow a route marked by tactile waypoints. To avoid participants using landmarks as orientation help, we did not include any visible objects except the ground. Repeating, chessboard-like textures on the ground allowed users to detect their movement visually. The application allowed the user to continuously turn left and right and move forward by holding the respective arrow keys pressed. A waypoint was reached, when the participant came closer than 40 Units.

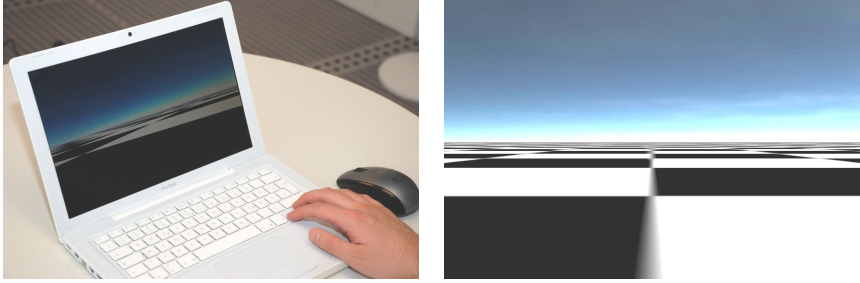


Fig. 4. Visual 3D presentation provided to the participants

All routes consisted of 5 sections with 3 waypoints each. The location of the waypoints was generated randomly with the constraint that each section had to be the same length (1000 Unit) and the same angle sum (287°). Thus, the ideal travel time for each route was the same, making the participants' performance comparable. Figure 4 shows the virtual environment and a user navigating along the route.

Each participant completed two routes, one using discrete and the other using interpolated presentation. The presentation method that was used first was randomly assigned to avoid systematic sequence effects. We measured the time the participants needed to complete each route to assess the participant's performance. We chose completion time as dependent variable, since besides distance covered by the participant it also takes situations into account where the user stands and turns around. Since movement speed and turning speed are both constant, this measure is an indicator for how accurate the participants stayed on the track as long as they do not halt without any reason.

Every participant had previously attended the first study. They were all introduced to the tactile belt and both presentation methods. The participants were told to complete the route as fast as possible and not to stop unnecessarily. Only, if the next waypoint was not in front of them, they should halt while turning. This should avoid the users walking away from waypoint, which would artificially degrade the measured performance. The experiment started, when the participants reached the first waypoint. Once they had completed the first route, the presentation method was changed and the second route started.

Results. We had no training session. Instead, we dedicated the first two sections of the route as training part. For the results we therefore only take the completion time of the last three sections of each route into account. We excluded two participant's results, because they experienced technical failures and thus struggled finding the waypoints. The completion time was significantly shorter with discrete presentation (65.59s, SE 3.05) compared to interpolated presentation (72.66s, SE 3.81). The presentation method had a large effect on the completion time ($t(12) = -2.67$, $p < .05$, $r = .72$). In general, participants walked straight towards the subsequent waypoint. We could not confirm the zigzag movement observed by van Erp et al. [13]. The most time got lost, if participants missed a waypoint and consequently had to turn around. In a few cases we observed

participants circling around a waypoint, missing it several times. This happened with both presentation methods.

4.3 Self Reports and Observations

After both experiments we handed out questionnaires, assessing the subjective impression of the participants about both display methods. The participants were asked to rate for each presentation method, how obtrusive it felt, how certain they were about the correctness of the directions they determined in the first task, how easy the determination directions in the first task was, and how easy it was to follow the route in the second task. Every aspect was rated on a five point Likert scale, ranging from zero (obtrusive/uncertain/difficult) to four (unobtrusive/certain/easy).

The ratings showed that during the first experiment the participants found discrete presentation (Mdn=3) significantly easier to interpret than interpolated presentation (Mdn=2). Discrete presentation had a medium effect on the difficulty (Wilcoxon signed-rank test: $T = 82, p < .05, r = 0.38$). There were no significant differences found between the presentation methods regarding the other three rating categories. Both presentation methods were found slightly comfortable in average (Mdn=3). Independent of the method, participants were slightly certain that the directions they had given were correct (Mdn=3), and found it slightly easy to follow the invisible route in the second experiment (Mdn=3).

During the evaluation several participants spontaneously mentioned that they were getting less sensitive to the vibration pulses. Consequently they found it harder to determine directions. In the first experiment, there were three cases where participants touched the belt to localize the tactile cues, since they were not able to determine their origin. Two participants even experienced a case, where they did not perceive the vibration anymore. After the first experiment, some participants indicated that determining directions was a very exhausting task and that especially interpolation is more difficult to interpret.

4.4 Discussion

The first experiment confirmed our assumption that the interpolated presentation method allows presenting directions more accurately compared to the discrete presentation method. This benefit came at the expense of slower reaction time. Despite the better accuracy, the second experiment showed that a route consisting of haptically presented waypoints is completed faster with discrete presentation. The questionnaires revealed no differences between the presentation methods except that the participants found it easier to interpret directions with discrete presentation in the first experiment.

The average accuracy of interpolated presentation found in the first experiment was similar to the accuracy we found with the same presentation method during earlier studies [4]. The accuracy of discrete direction presentation was

worse than anticipated. As explained in section 3 we expected an inherent inaccuracy of 15° for the discrete presentation. The actual result was 19.43° . The difference between those values can serve as an indicator for variance decreasing the accuracy, such as the difficulty of mapping the physical experience to a visual circle or cases when participants were not able to determine which vibrator was activated.

The participants' impression that interpolated feedback is more difficult to process is backed up by the in average longer reaction time measured in the first experiment. Previous work showed that processing tactile feedback can be successfully trained [6]. However, we could not find a significant difference caused by previous experience with the tactile belt. We therefore suggest that if processing interpolated feedback can be trained, it requires more practice.

The longer reaction time might also have been a reason for discrete presentation resulting into faster completion of the route, in the second experiment. We suspect that slight changes were noticed later due to the longer processing time of interpolated feedback and due to vibration insensitiveness, since the feedback was almost always perceived from the front.

Our assumption that interpolated presentation is perceived as less obtrusive could not be confirmed. However, due to the nature of the conducted experiments those abrupt changes in the output were perceived using both presentation techniques. In the real world the relative direction of objects, such as waypoints and POIs, do not change abruptly but continuously. Sudden changes in the display's tactile output would not occur with interpolated presentation. Thus, we assume to obtain different results for real world tasks.

5 Conclusion and Future Work

In this paper we presented a presentation method for tactile belts that displays any direction in the horizontal plane using six vibrators only. The developed presentation method displays a direction by interpolating the intensity of two adjacent factors. In two experiments this presentation method was compared to the discrete presentation method used by other belt type tactile displays. The experiments showed that interpolated presentation is more accurate than discrete presentation for the developed tactile belt with six vibrators. However, it was also found that interpolated presentation is more exhausting and takes longer to be interpreted. Despite the increased accuracy, interpolated presentation performed worse in a waypoint finding task compared to discrete presentation.

In our future work, we will use the data of the presented study to refine the interpolation. A promising approach is to replace the used linear interpolation function by a more sophisticated one. The extinction of the tactile sensation also has to be considered. Additionally, we plan to experiment with encoding additional information through other parameters of the tactile output. In particular, we are interested in using tactile displays for the presentation of localized objects, such as POIs, landmarks, or persons.

References

1. Brewster, S., Brown, L.M.: Tactons: structured tactile messages for non-visual information display. In: Proc. of the Australasian conference on user interface (2004)
2. Brewster, S., Wall, S., Brown, L., Hoggan, E.: Tactile Displays. In: The Engineering Handbook on Smart Technology for Aging, Disability and Independence. Computer Engineering Series, John Wiley & Sons, Chichester (2008)
3. Henze, N., Heuten, W., Boll, S.: Non-intrusive somatosensory navigation support for blind pedestrians. In: Proc. of Eurohaptics 2006 (2006)
4. Heuten, W., Henze, N., Boll, S., Pielot, M.: Tactile wayfinder: A non-visual support system for wayfinding. In: Proc. of NordiCHI (2008)
5. Lindeman, R.W., Sibert, J.L., Mendez-Mendez, E., Patil, S., Phifer, D.: Effectiveness of directional vibrotactile cuing on a building-clearing task. In: Proc. of the conference on human factors in computing systems (2005)
6. Nagel, S., Carl, C., Kringe, T., Martin, R., Konig, P.: Beyond sensory substitution. Learning the sixth sense. *Journal of Neural Engineering* 2 (2005)
7. Pielot, M., Henze, N., Boll, S.: FriendSense: Sensing your Social Net at Night. In: Workshop Night and Darkness: Interaction after Dark in conjunction with CHI 2008 (2008)
8. Rupert, A.: Tactile situation awareness system: Proprioceptive prostheses for sensory deficiencies. *Aviation, Space and Environmental Medicine* 71, 92–99 (2006)
9. Tan, H.Z., Pentland, A.: Tactual displays for wearable computing. In: Proc. of the International Symposium on Wearable Computers (1997)
10. Tsukada, K., Yasumura, M.: Activebelt: Belt-type wearable tactile display for directional navigation. In: Proc. of the Conference on Ubiquitous Computing (2004)
11. Van Erp, J.B.F.: Guidelines for the use of vibro-tactile displays in human computer interaction. In: Proc. of Eurohaptics 2002 (2002)
12. Van Erp, J.B.F., Groen, E.L., Bos, J.E.: A tactile cockpit instrument supports the control of self-motion during spatial disorientation. *Human Factors: The Journal of the Human Factors and Ergonomics Society* 48, 219–228 (2006)
13. Van Erp, J.B.F., Van Veen, H.A.H.C., Jansen, C., Dobbins, T.: Waypoint navigation with a vibrotactile waist belt. *ACM Transactions on Applied Perception* 2, 106–117 (2005)

Supporting Collaboration between Visually Impaired and Sighted Children in a Multimodal Learning Environment

Erika Tanhua-Piiroinen¹, Virpi Pasto¹, Roope Raisamo¹, and Eva-Lotta Sallnäs²

¹TAUCHI, Department of Computer Sciences
33014 University of Tampere
Finland

{erika.tanhua-piiroinen, virpi.pasto, roope.raisamo}@cs.uta.fi

²KTH - Royal Institute of Technology
SE-100 44 Stockholm, Sweden
evalotta@csc.kth.se

Abstract. Visually impaired pupils are a group that teachers need to pay attention to especially when planning group work. The need for supporting collaboration between visually impaired and sighted people has been pointed out but still there are few evaluations on that. In this paper two studies are described concerning collaboration support for visually impaired and sighted children in a multimodal learning environment. Based on the results of the first study where two children used a multimodal single-user Space application together, the application was improved to better support collaboration. This prototype was then evaluated. According to the results it is worthwhile to provide individual input devices for all the participants in the group. For helping the pupils to achieve a common ground it is also important to provide sufficient support for all senses in a multimodal environment and to take care of the feedback about the haptic status of the environment also for the sighted participants.

Keywords: Collaboration, visually impaired children, multimodal interaction, haptics.

1 Introduction

Children with special needs are nowadays often integrated in ordinary classes which make the learning situation more challenging both to the children and to the teacher. One of those special groups that teachers need to pay attention to is visually impaired children. These pupils may have a lot of different assistive resources in the classroom, like personal assistants, Braille books, relief pictures or maps and voice synthesizers. But still they can feel as outsiders during the lessons and also during the breaks. This can especially happen when pupils are doing group work together [11]. The special tools that visually impaired pupils have in school are usually not designed for group work but for individual work. The need for supporting collaboration between visually impaired and sighted people has been pointed out before and recommendations for how a graphical user interface should be presented in other ways than visually have been suggested [2, 4]. However, there are few evaluations on multimodal support for group work in a school context including visually impaired and sighted pupils.

Visually impaired and sighted pupils do group work together on various topics in the school today [11]. A problem is, however, that during group work visually impaired pupils partly work in a separate work process rather than included in the main group work. This happens in spite of the good intention to include everyone in the work. Collaboration depends on an unrestrained dialogue and on the fact that people working together can maintain a common ground of the context of joint action including the features of the workspace and the continuous status of the work process. Some of the basic features of face-to-face conversation are: a shared physical environment, ability to see and hear each other, ability to perceive each other's actions and that people can produce and receive communication at once and simultaneously [1]. In the case of group work between visually impaired and sighted pupils in the school today, a shared understanding can not always be easily maintained. The respective work tools of the pupils do not allow shared mutual access to information or perception of others' actions. Thus both the work space and the work process have an influence on this lack of shared understanding.

In recent years different multimodal interaction models have been developed and investigated for visually impaired children. User interfaces have been implemented for different haptic devices, for example, game controllers and the SensAble Phantom device. Usability tests have been conducted investigating the benefits of haptic feedback in different applications [5, 7, 9, 13]. A Space application has been designed to support children's exploratory learning among astronomical phenomena [8].

Investigations of the Space application and also most of the previous studies have focused on single user situations. During the last three years a multimodal learning environment for inclusion of visually impaired people has been designed in a European project MICOLE (Multimodal Collaboration Environment for Inclusion of Visually Impaired Children). The focus has been collaboration between visually impaired and sighted children (see [6]). As a part of this research project the Space application has been investigated in a collaborative setting and based on the results of that study it has been further developed to support collaboration and then tested again. In this article we describe the results on children's collaboration in these two studies.

2 Systems and Applications Tested

2.1 Apparatus

A SensAble Phantom Desktop device [12] with a stylus that is used like a pen was used to make it possible for the child to feel the contents of the application. A Magellan space mouse (Fig. 1 in the middle) was another input device used in both studies. In the Solar System application an ordinary computer mouse and a keyboard were also used. With the computer mouse a child can guide the Phantom user by haptic feedback to a certain target. The cursor of the mouse is a red cross and moving this cross and pushing the left button of the mouse, the Phantom stylus will move to the place marked with the cross. The visual feedback was provided with 17" LCD displays. Stereo sound was provided with loudspeakers.

In the first study the Phantom was placed in front of the visually impaired child and one LCD display was placed in front of the sighted child (Fig. 1, left).

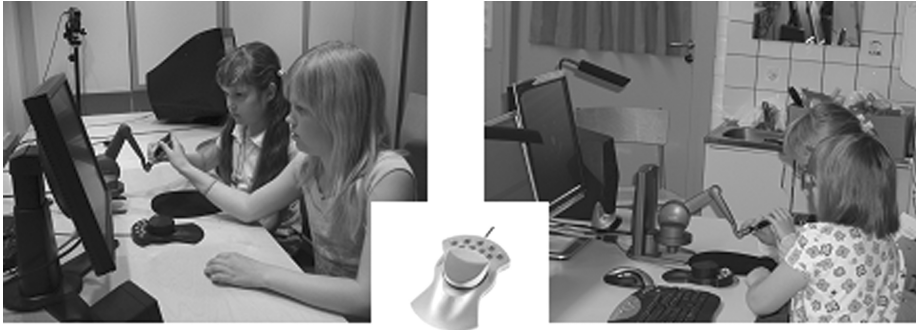


Fig. 1. A visually impaired and a sighted child using the Phantom together in the first study (left) and in the second study (right). The Magellan Space Mouse [14] in the middle.

The test setup of the second study (Fig. 1, right) consisted of two 17" LCD displays. In the application used in this study the small buttons of the Magellan space mouse were utilized as well as the big one.

2.2 The Space Application

The Space application [9] was designed to support children's explorative learning. Four micro worlds were used in the study: the Solar System, the Earth, the Earth's Orbit and the Earth's Internal Layers (Fig. 2). There is also a menu for navigating to each micro world. The Earth can be touched with the stylus and distinct surfaces like oceans and the ground feel differently. It is also possible to rotate the Earth. In the Solar System the orbits of the planets can be followed with Phantom stylus. When a certain planet is found it is possible to listen to some extra information about it by pushing the stylus at the planet. In the Earth's Orbit environment the Earth can be moved with the stylus along its orbit and the application tells the user which season there is in Finland at every position. Finally, in the Earth's Internal Layers micro world different compositions of the layers are illustrated as a cross-section where they can be touched and the descriptions about layers can be heard. The user navigates back to the menu from every micro world by pushing the big handle on the space mouse.

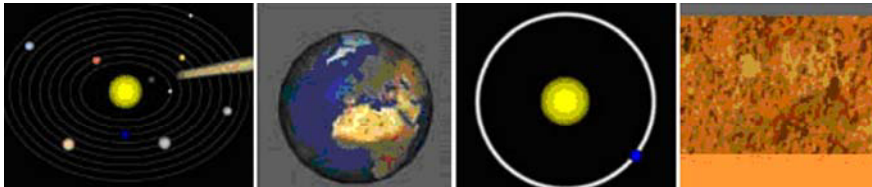


Fig. 2. The micro worlds of the Space application: The Solar System, the Earth, the Earth's Orbit and the Earth's Internal Layers

2.3 The Solar System

The Solar System application is a part of the Space application. It was developed in order to be used by visually impaired and sighted children together in a collaborative situation. The new features make it possible to make notes and to guide the Phantom user by the computer mouse.

The application makes use of two screens: The Solar system user interface in one and the Notebook view in another one (Fig. 3). With the Solar System the user can explore the orbits of the planets with the Phantom stylus and the speech synthesizer tells which planet's orbit is in question. When finding a planet the user can explore its surface by pushing the planet with the stylus. These planet surfaces are new micro worlds that have been designed for this application. The Sun can also be explored.

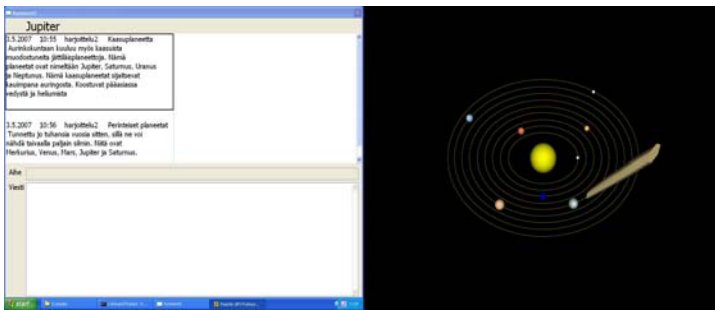


Fig. 3. Two screens of the Solar System

On the surface of the object it is possible to feel the shapes and textures of it and listen to some extra information by pushing the little button on the stylus. The other screen shows a notebook view consisting of three fields and the title relevant to the context. The title can only be read but the contents of the other fields can also be heard. In the uppermost field all the previous notes written in this certain context can be seen. The other two fields are used for writing a note: a title field and a content field. An ordinary keyboard is in use for typing but the Magellan space mouse is used for manipulating the notes. In this way both children have equal possibilities to use these functions.

3 Evaluation Settings and Procedures

The laboratory where the first study took place consisted of two rooms. In the testing room the children operated the application and the device with the aid of an assistant who was a nursery school teacher. In the observation room two test organizers were observing the test situation through a one-way glass. Also, the parents had an opportunity to follow the test from that room.

In the second evaluation at schools, the testing was organized in separated rooms where only the participants of the test and the test organizers were present.

The testing procedure for tests with visually impaired children has been developed in our previous research projects [5]. Before the evaluations we conducted pilot tests to investigate both the usability of the application and the adequacy of our testing procedure in collaborative settings.

The test began with a casual start-up interview concerning the children's experiences on computers and applications and possible earlier use of haptic devices. The children were then interviewed about group works at school. After that, the children were shortly briefed on the contents of the application. Then they were taught to use it by means of real world models and a plastic stick, which was of similar shape to the one of the Phantom device. Both children were also taught to hold the Phantom stylus properly and how to use that device.

At first the children used and explored the application freely and after that they were asked to complete different tasks utilizing the application. The tasks were simple as our aim was not to test children's learning processes this time but to investigate the collaborative use of the multimodal application in hand. The tasks were read aloud and given also on paper to the children. When the tasks were completed the children were finally interviewed about their experiences of using the application together.

The analyses were made of the videotaped interaction and the post test interviews. Two evaluators analyzed the children's activity during the group work, regarding which one of them (the visually impaired or sighted) was the most dominant, if any, and which one of them took the most often the initiative. The number of observed actions by the children and their verbal and non-verbal expressions regarding these dimensions were also noted. Finally, the ways in which the children helped each other and the help needed from the test assistant were analyzed.

4 The First Study

4.1 Participants

Four pairs of children (one blind or visually impaired child and one sighted child) were invited to our laboratory to test the Space application. The participants were 9–12 years old, two of them were girls (one pair) and the other six were boys. Three of the visually impaired children were blind. We asked the visually impaired children to ask a sighted friend to join them in the test, and as a result the children's relationships were different in every pair: schoolmates, cousins, siblings and twins.

4.2 Tasks

In this first study the children solved three tasks one by one and told the answers verbally to the test assistant. The questions were the following: 1) Which are the reasons for the fact that the Earth is the only planet with life in our solar system? 2) Which continents are not mentioned in the part of the application that contains information about the oceans and continents of the Earth? 3) Why is it not possible in the real life to dig a tunnel deep enough to come out from the other side of the Earth?

4.3 Results

Task performance. All the pairs managed to finish all the tasks. One of the pairs (twins) succeeded very well and completed the tasks without assistance. The other three pairs needed either verbal guidance or hands-on help in using the stylus.

Activity. Both the blind or visually impaired and the sighted children were active during the collaboration. They completed the given tasks together and not individually. Although the visually impaired child was the “main user” of the Phantom stylus, they also changed roles during the tasks when needed.

Commitment. The children concentrated quite well in the tasks. However, most of them had already knowledge about the application area and maybe this reduced a bit their motivation in completing the tasks. In some cases we noticed additional actions (outside of the tasks) by the child who was not handling the Phantom stylus. This happened when the child did not have so much to do at that time with the application.

Dominance. Children’s joint use of the stylus occurred mostly by the sighted child as an initiator. The reason was probably that the blind or visually impaired child was the main user of the stylus. In one test session the blind child held the stylus all the time with both of his hands. This made it problematic to use the stylus together with the other child.

In those cases where children were not so much used to work together the blind or visually impaired child seemed to dominate more. This was especially evident when the visually impaired children were active and enthusiastic in doing the tasks and the sighted children (also enthusiastically) grabbed the stylus but released it immediately when the visually impaired did not support that kind of joint interaction. In one case the blind child pushed the other’s hand away from the stylus, but there were however no signs of actual dominance; this pair worked seamlessly together as if they had a joint work plan. Neither of them needed to dominate. Both of the children were concentrating well on the common task.

4.4 Lessons Learned

The child using the Phantom stylus often dominated in collaboration. Usually this child was the visually impaired one, who was seated in front of the Phantom device. In this case the role of the sighted child was more like an assistant than a participant.

Although the sighted child evidently wanted to grasp the stylus, he or she hesitated to do that or the visually impaired child even pushed the other’s hand away from the stylus. One solution to avoid this situation could be to offer each child an own interaction device.

There could be much more visual information for the sighted child as the haptic feedback is mainly intended for the visually impaired child. This could increase the commitment of the sighted child to the learning situation. The sighted child could e.g. get visual feedback of the haptic information that the visually impaired gets through the Phantom. Then, the children would get a better shared understanding of the workspace that would help both children to discuss information, and to guide each other during navigation and exploration in the application.

On the other hand, if both children had one haptic device each that would solve that problem. The fact that the children spontaneously held on to one stylus shows that they both had the need to get the haptic feedback. But holding on to the same stylus, one child having a hand on top of another's hand, is not good ergonomically.

Some pairs in our tests wanted to write down the answers of the tasks. They asked to have paper and pencil, which were not included in our test set-up. Paper and pencil are common tools for children in the school environment and should thus be available in the test situation, too. On the other hand the application could offer the opportunity to make notes.

Testing the application in a laboratory and not in a real learning situation in a classroom does not give the best insight into how it supports children's collaborative learning. Thus, we decided to test it in real school contexts.

5 The Second Study

According to the results of the first evaluation we further developed the application. The resulting Solar System application was introduced in Section 2.3. The children used the application at the schools and all the sessions were videotaped. As the application had been changed and new collaborative tools had been designed, the evaluation was focused on the usability of those new parts. Apart from that, the collaborative issues were assessed in the same way as in the first evaluation. When designing the test tasks we planned them with a teacher of geography and she was also observing the first pilot test.

5.1 Participants

Three visually impaired and three sighted pupils participated in testing which took place in three schools. These children were not the same as in the first study. The ages of the children were between 7 and 9 years. One of the visually impaired children was blind; the other two could use their sight a little.

We also experimented with one visually impaired child and his sighted partner both of them having an additional hearing defect. However, the hearing defect has an effect on children's conceptual understanding, and thus the testing had to be specifically tailored for them. In the analysis of the results this pair has been left out because of the different testing procedure.

5.2 Tasks

The tasks were the following: 1) Locate Finland on the globe. 2) How long does it take a shaft of sunlight to reach the surface of the Earth after its departure from the surface of the Sun? 3) Which are the two stone planets of our solar system that have no moons? 4) Why is it impossible for people to live on the other planets? Examine three planets of your choice and write down at least one reason for every planet as an answer. (Before the fourth task the test assistant discussed with the children about prerequisites of life, and how is it possible to live on the Earth.)

The answers were written as notes in the application. Children were advised to use the application together and also to think aloud. They could be helped in making up the titles for their own notes which proved to be quite difficult for the children in this age according to the pilot tests.

5.3 Results

Task performance. All three pairs of children without hearing defect managed to use the application and they finished all the tasks. However, all of them needed help in some tasks and the most common type of assisting was motivating. The possibility to make notes was a successful feature and all the children in pilot tests as well as in the main tests liked to hear their own comments read by the synthesizer. Also the possibility to guide was greeted with satisfaction.

Especially in the first test task where the children were asked to find Finland on the globe, it was observed that exact locations were difficult to find even when guided with the mouse.

Dominance. In the first study we found that the child using the Phantom device seemed to dominate in collaboration. Now there were more interaction possibilities in the application. However, when one partner was using the Phantom the other one did not have much to do, and when the other one was writing the note the Phantom user got a bit bored. Thus, the dominance seemed to be linked to the use of the input devices. It was technically possible to use the Phantom and to write a note simultaneously, but in our procedure (with the tasks) that was not very natural. During the sections when only one child was actively operating the stylus or writing a note they still collaborated verbally.

Common ground. The children managed to complete the tasks well together and the discussion about the task on hand helped them through the tasks. The structure of the application (two screens with different functions each) might have an influence on the work flow: the children took turns in doing actions and sometimes this caused a break in collaboration. The division of work was a challenge in writing notes in the collaborative setting. The saving function didn't give any feedback through sound or haptics. The saved text just disappeared and came visible in the list of notes. So the visually impaired child didn't realize that the text had been saved.

Guiding. The children assisted each other but all the guidance was given by the sighted partners. Most of time they guided by telling the other one in which direction to move or where some object could be found, but they also used the possibility to guide the other child with the mouse. One feature in the way the Phantom worked confused the children: when the stylus is positioned in peripheral areas of the haptic field (an edge), where two surfaces are very close to each other, the stylus starts to shake a bit. This made children comment like "hey, what's that, what's happening?" However, in one case the sighted child guided the Phantom user away from the shaking edge by the mouse, which was an indication of children's reciprocal helping in an extraordinary situation.

6 Discussion and Conclusions

The exact locations were difficult to find even when guided with the mouse. Additional investigations are still needed to find out if this is caused by the Phantom, the mouse or maybe by both. One solution could be to replace the mouse with another Phantom device, like it was done by Sallnäs et al. [10] or McGookin and Brewster [3]. As the mouse locates targets two-dimensionally and the Phantom supports three dimensions, this improvement could help in guiding to also very strict points.

Only the guiding with the mouse occurred always simultaneously and the notes could be written both sequentially or in parallel with the exploring. So the awareness of the situation of the partner does not have as big an influence on completing the tasks together as in the case where the tasks have to be done simultaneously all the time. As the children collaborated quite well by discussing about what was happening, the occasional lack of awareness or common ground on the other hand could bring along active collaboration. If the children had been working all the time simultaneously, they would have possibly missed some fruitful discussions.

Based on our findings we recommend the following. Some kind of input devices should be provided to all children that participate in the group work so that all children can interact with and explore the workspace. This can reduce the device dominance phenomenon. Furthermore, sufficient visual feedback should be provided of the haptic information for the sighted child, if another haptic device is not available for her or him. Finally, ergonomic issues are very important to take into account when designing a learning environment, especially when many devices are to be included there.

Acknowledgments. The project MICOLE (IST-2003-511592 STP) was funded by the European Commission. This work was also partially supported by the Academy of Finland (project 114079). We thank the staff who planned and implemented the applications and our partners in Europe. We also thank Irma Sommers for her contribution in task planning and Erno Mäkinen and Rami Saarinen for their valuable comments on this article.

References

1. Clark, H.H.: Using language. Cambridge University Press, Cambridge (1996)
2. Edwards, K., Mynatt, E., Stockton, K.: Access to graphical interfaces for blind users. *Interactions* 2, 154–167 (1995)
3. McGookin, D., Brewster, S.: An initial investigation into non-visual computer supported collaboration. In: *CHI 2007 Extended Abstracts on Human Factors in Computing Systems* (San Jose, CA, USA, April 28 - May 03, 2007), pp. 2573–2578. ACM Press, New York (2007)
4. Mynatt, E., Weber, G.: Nonvisual presentation of graphical user interfaces. In: *Proceedings of CHI 1994*, pp. 166–172. ACM Press, New York (1994)
5. Raisamo, R., Hippula, A., Patomäki, S., Tuominen, E., Pasto, V., Hasu, M.: Testing usability of multimodal applications with visually impaired children. *IEEE Multimedia* 13(3), 70–76 (2006)

6. Rasmus-Gröhn, K., Magnusson, C., Efring, H.: AHEAD – Audio-Haptic Drawing Editor and Explorer for Education. In: HAVE 2007 - IEEE International Workshop on Haptic Audio Visual Environments and their Applications, pp. 62–66. IEEE Press, Los Alamitos (2007)
7. Rasmus-Gröhn, K., Magnusson, C., Efring, H.: User Evaluations of a Virtual Haptic-Audio Line Drawing Prototype. In: McGookin, D., Brewster, S. (eds.) HAID 2006. LNCS, vol. 4129, pp. 81–91. Springer, Heidelberg (2006)
8. Saarinen, R., Järvi, J., Raisamo, R., Salo, J.: Agent-based architecture for implementing multimodal learning environments for visually impaired children. In: Proceedings of ICMI 2005, the 7th international Conference on Multimodal interfaces (Toronto, Italy, October 04 - 06, 2005), pp. 309–316. ACM Press, New York (2005)
9. Saarinen, R., Järvi, J., Raisamo, R., Tuominen, E., Kangassalo, M., Peltola, K., Salo, J.: Supporting visually impaired children with software agents in a multimodal learning environment. *Virtual Reality* 9(2-3), 108–117 (2006)
10. Sallnäs, E.-L., Bjerstedt-Blom, K., Winberg, F., Severinson Eklundh, K.: Navigation and Control in Haptic Applications Shared by Blind and Sighted Users. In: McGookin, D., Brewster, S.A. (eds.) HAID 2006. LNCS, vol. 4129, pp. 68–80. Springer, Heidelberg (2006)
11. Sallnäs, E.-L., Crossan, A., Archambault, D., Tuominen, E., Stoeger, B.: Report on development of collaborative tools - User requirements study and design of collaboration support. Deliverable D8, MICOLE project (2005), <http://micole.cs.uta.fi/>
12. SensAble PHANTOM. Products & Services, <http://www.sensable.com/haptic-phantom-desktop.htm>
13. Zijp-Rouzier, S., Petit, E.: Teaching geometry to visually impaired pupils using haptics and sound. In: Proceedings of HCI 2005, Universal Access in Human-Computer Interaction conference (2005)
14. HP Spacemouse Plus, <http://www.dooyoo.co.uk/mice-trackballs/hp-spacemouse-plus/>

Perceptually Informed Roles for Haptic Feedback in Expressive Music Controllers

Ricardo Pedrosa and Karon MacLean

Department of Computer Science, The University of British Columbia,
Vancouver, BC. V6T 1Z4, Canada
{rpedrosa, maclean}@cs.ubc.ca

Abstract. In this paper, we propose a methodology for systematically integrating haptic feedback with a co-developed gesture interface for a computer-based music instrument. The primary goal of this research is to achieve an increased understanding of how different sub-modalities of haptic feedback should be combined to support both controllability and comfort in expressive interfaces of this type. We theorize that when including haptic feedback in an instrument, force and vibrotactile feedback could be beneficially designed *individually* and then fine-tuned when mixed in the final design.

Keywords: Haptics, gesture interface, perception.

1 Introduction

Today's electronic music instruments often take the shape of more traditional instruments (like keyboards, electric guitars or wind controllers), adding digital functionality while retaining both the strengths (typically including high controllability) and usability flaws of their models. Conversely, contemporary computer-based music interfaces depart so greatly from traditional acoustic instruments that they can be hard to recognize as instruments at all. Some novel approaches, such as gesture controllers, are quite expressive and easier to learn than acoustic instruments, but they also tend to be less controllable and do not capture the characteristic "*feel*" of a music instrument.

The research proposed here aims to identify at least one path by which the *feeling* of using a music instrument can be captured by gestural computer-music interfaces, without compromising the expressiveness and ease-of-use they already exhibit. Specifically, we wish to discover how to most beneficially combine sub-modalities of haptic feedback (namely, force and tactile feedback) to simultaneously support controllability, comfort and expressivity in interfaces such as computer-based music instruments.

The three perceptual modalities involved in musical performance (auditory, visual and haptic) are deeply implicated in the construction of music instruments. For instance: to fulfill the main objective of enabling a distinctive and enjoyable sound, a luthier (a craftsman who makes stringed instruments) must base his/her design upon the basic movements with which the musician will control the sound-generating mechanism. In the construction stage, the luthier chooses materials to fulfill one objective: the instrument's body should amplify the sound produced by the vibration of the sound-generating mechanism and, most importantly, must resonate in a way

that reinforces certain frequencies of the sound spectrum, thus defining the timbre for the particular instrument. This vibration is felt as well as heard. Thus, control movements provide both critical haptic communication with the instrument, and visual communication with other musicians and the audience.

1.1 Haptics as a Possible Design Approach

Despite being the locus of physical interaction between musician-instrument, the haptic channel has received little attention since new computer interfaces for music began to appear. This is because designing these interfaces takes on a different path away from the luthier's approach of selecting parts and configuring the instrument to amplify the sound generated. The first inclusion of haptic feedback in electronic music interfaces came as a result of musicians' requests to incorporate the mechanical feeling of an acoustic piano into electronic keyboards. As opposed to the case of acoustic instruments where the haptic feedback is a physical artifact of the sound-generating mechanism, in electronic instruments it must be deliberately added. This is especially true in the case of those interfaces that depart more drastically from the traditional concept of a music instrument, yet have the potential for the highest levels of expressivity and usability.

1.2 Related Work

1.2.1 Gesture in Music Performance

One of the most important features in music performances is the communication between the artist and the audience. Not only is the audience able to perceive the effort and ability of the performer through his/her gestures, but also "the musician is able to communicate to the public the inner meanings and movements of the music being played" [13] through both grounded and un-grounded gestures. In grounded gestures, forces are exchanged with the instrument and enforce the musician's skills by providing references to signal for instance, the completion of an action (e.g. the hammer release after pressing a piano key). Ungrounded gestures are performed in air, e.g. by a conductor's baton or by the musician's body movements, and are most responsible for conveying the music's "inner meaning". A great challenge in designing a computer-based instrument thereby lies in deploying an interface that allows at least as much gestural freedom as does a given traditional instrument. The traditional computer interface of a keyboard and a mouse is simply not enough for these purposes.

To solve this problem, one approach (termed "open air" or "immersive" controllers) employs remote gestural sensing, mapping each gesture to an acoustic event or a control method. However, despite being considered the most intuitive interfaces and the ones that could be more easily learned by a general public, these gesture interfaces do not capture the intimate physical interaction on which acoustic instruments are built. Gesture controllers demand a high degree of proprioception (the sensation of movement and spatial orientation arising from musculoskeletal sensors) and egolocation (awareness of one's overall position within a defined space); yet further rely on visual and auditory feedback to achieve movement accuracy. Musicians trying to master these interfaces must then develop the same body control of a dancer, a process that requires a totally different set of skills and continual concentration.

1.2.2 Musician-Instrument Feedback Control

Music instruments are complex mechanical resonators: in order to produce sounds they must vibrate. It is known that these vibrations are used to help the performer in tasks ranging from determining when the note is settled and stable in the instrument [7] to tuning the instrument to a neighbor based on vibratory harmonics transmitted through air or floor [1].

Gillespie [11] provided a broad definition of a music instrument as: “a device which transforms mechanical energy (especially that gathered from a human operator) into acoustical energy”. Thus, the musician-instrument relationship could be represented as a feedback control system [12] [14] with the control loop closed via both the auditory and haptic channels. The primary feedback from instrument to musician occurs through the air in the form of sound waves. However, the grounded mechanical contact between the musician and the instrument (e.g. fingertips, hands or mouth) acts as an additional bidirectional channel through which the musician sends a continuous command signal in the form of mechanical energy, and receives haptically a sense of the sound generated and the status of the sound control mechanism.

Repp’s [16] and Finney’s [9] results show that after years of training, the musician is able to perform a form of anticipatory control over the instrument where she/he anticipates an instrument’s response to a given manipulation.

In a thorough study of vibrations in four traditional stringed instruments, Askenfelt and Jansson [1] provide enough evidence to assert that mechanical vibrations occurring in music instruments are powerful enough to be felt by the musician during regular performance, and that these vibrations are not limited to the parts of the instrument designed to radiate sound. They also point out that regardless of the type of instrument, it could be assumed that the kinesthetic finger forces offer more guidance for timing purposes than the vibrotactile stimuli supplied by instruments’ vibrations.

Taken together, these works strongly suggest that the haptic submodalities of force and vibrotactile feedback have distinct functions in music performance, playing important roles in both fine-tuning performance and defining instrument status. Forces exerted by the musician over the instrument are the main channel through which the sound output is controlled: depending on how strong a string is plucked, a key is hit, a drum is banged or a mouth is closed on a mouth piece will determine the type of sound generated. The development of low-level sensorimotor programs through practice and training [12] depends heavily on the musician’s apprehension of the relationship between force exerted and sound produced. Vibrations from the instrument’s sound generation engine close this local, mechanical loop, while reinforcing the feeling of using a music instrument -- or, in the words of Askenfelt and Jansson, a “resonating and responding object”.

1.2.3 Haptic Feedback in Electronic Instruments

Designers include haptic feedback in computer music instruments for a variety of reasons. It tends to make new interfaces more attractive to the performers by fostering (among other attributes) expressivity or controllability. In some cases it is the only way to control some recent sound synthesis algorithms [8][15]. In others, the sense of controllability is sometimes enhanced by haptic realism in the mimicking of the feel

and dynamic behavior of a traditional music instrument [10][5]. Conversely, O’Modhrain [14] demonstrated benefits of haptic feedback in instruments that traditionally provide no haptic cues.

With respect to open air gesture controllers, several approaches have addressed issues of inconsistent control and unrepeatability. Force feedback in the form of shape memory alloys [2] and pneumatic devices [4] has been added; Chu [6] proposed a haptic MIDI controller to localize sound in space with tactile-audio signals to both hands. Rován and Hayward [17] enhanced an open-air glove controller with tactile actuators on the performer’s hands or feet, to explore the perceptual attributes of a simple vibrotactile vocabulary synthesized in response to a gesture.

2 Proposed Methodology

2.1 Proposition

While designing haptic feedback in computer interfaces, the general path has been to either (a) use only one of the haptic channels and push to the limit the amount of information conveyed through it or (b) use various channels and design the interaction based on a predefined fixed schema where the feedback delivered is set and adjusted taking into account technological factors rather than perceptual issues.

We theorize that when including haptic feedback in any computer interface, force and vibrotactile feedback could be beneficially designed individually and then fine-tuned when mixed in the final design. Extra cues to be supplied to the user should likewise be designed and mixed in each haptic submodality before the final blend occurs. For music interfaces, these extra cues (ie. feedback from other instruments or performers, cues on rhythm or tempo, etc) might not play a direct role in enforcing controllability, but could reinforce the perception of more complex music parameters like contour (ie. the shape of a melody when musical interval size is ignored).

To put this theory into practice there are (at least) two conditions that must be satisfied: (a) it is strictly necessary to have a thorough knowledge of the context of the interface (here, the context is that of an expressive music interface); and (b) it is necessary to know the role of each of the haptic channels in that situation. With that knowledge and a design strategy that takes it into consideration, the designer could achieve a higher knowledge transfer from related non-digital contexts (e.g. acoustic interfaces). This will increase the usability for both expert users (for them it is like using the tools they are used to) and beginners (skills developed in related contexts like dance or sports where some sort of continuous/harmonic behavior is needed could be applied here).

On the other hand, this “divide and conquer” strategy has a drawback. In expressive interfaces, the feeling of comfort is as valued as the feeling of being in control. There are several ways a feeling of discomfort may arise, one being undesired system feedback that interferes with the user’s goals when these are not in accord. Most other ways are due to an overloaded perceptual channel or to interaction or masking between information presented through different channels. For instance, when a supplied stimulus is too strong or maintained for too long it could numb the receptors to a point that consequent stimuli are not felt; if occurring too close in time, information is lost. Tactile

and kinesthetic signals might be more vulnerable to mutual masking than, say, tactile and visual cues. Designers could take advantage of masking to avoid overloading: by presenting some important information as a masking signal only this one will be felt.

2.2 Paradigm and Platform

The literature review shows that gesture interfaces are considered among the best option to address computer based music instruments because of their inherent expressivity and intuitiveness but at the same time are the ones where the intimate relationship between musicians and instruments become harder to recreate.

To test our assumptions, we propose a research path focused on a case study of a grounded gesture interface capable of mapping in 3D the movement of a performer's hand, at the same time that it provides both force and vibrotactile feedback to the hand. This paradigm should have sufficient richness to control a set of expressive parameters in an arbitrarily selected music performance, and will provide the means of evaluating our metrics of comfort and controllability.

As our interest is concentrated on designing and evaluating the effects of haptic feedback in relation to expressive applications rather than the design of the gestural / haptic interface itself, our approach will take as a base a commercially available device that provides both a 3D mapping mechanism and haptic feedback through the force channel (a PHANTOM Premium 1.5 from SensAble Technologies). We will design and build an add-on to the original interface to include the desired vibrotactile feedback that the original device does not provide.

This platform will also allow us to test our primary hypothesis (the separable design of haptic feedback channels based on their perceptual roles in a given task context) in a best-case scenario, in that the platform, while not able to reproduce real-world fidelity, does deliver relatively high-quality and high-DF (degrees of freedom) haptic feedback. This approach is preferred to an approach of commodity-level hardware, because we will be able to (a) eliminate feedback combinations that do not work by focusing on perceptual issues rather than technological ones; and (b) demonstrate that those combinations that work can be deployed using available technology. Furthermore, (c) satisfactory results obtained through this research could be narrowed down to the optimum point of performance/cost through a degradation study where the top-of-the-line characteristics of this interface are deliberately reduced to determine the values necessary to produce required performance ratings, as in [3].

2.3 Stages

The proposed research will follow the path described in Figure 1, with three design and two mixing stages. Most stages include both development and user evaluation.

2.3.1 Development of Semantic Base for Application Gestures

The semantic base that we aim to gather will consist of meaningful gestures that could suit our purpose of controlling a music interface. The movements' syntax (the structure by which a performer combines each of the control actions, or gestures, to achieve certain aesthetic results) will be left to the performer within the interface constraints.

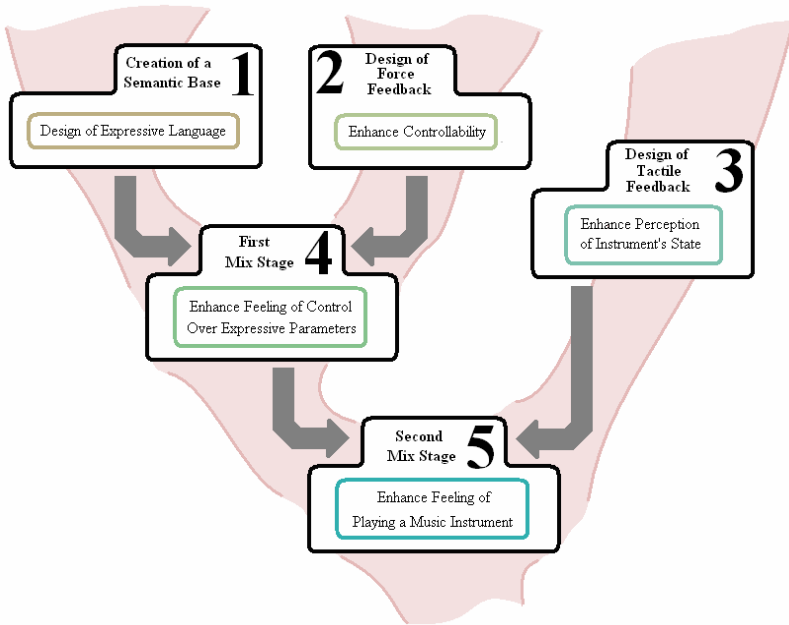


Fig. 1. Proposed research stages. The methodology is structured as a roadmap where the different stages converge at different levels and time frames to fulfill different goals.

Our approach is to watch a broad set of users using the interface to mimic the generation of sounds and extract from those performances a set of gestures suitable for our purpose. This observational study will allow us to determine if there is a generic approach for several control parameters (ie. pitch, intensity, etc.) and at the same time will give us an extended set of gestures over the ones we might have predicted in advance.

As we are building an expressive interface, we are most interested in the meaning of a gesture and how it should be mapped to control a certain music parameter, as opposed to the exact free gesture trajectory. At this stage, the semantic base will be constrained by the PHANTOM's kinematics and workspace, but not by active forces. In general, an ungrounded gesture can not be repeated in the same way when some impedance is added by using a grounded interface. By focusing on the meaning of the free gesture, we anticipate that this semantic base will be applicable to the broadest spectrum of actual gesture interfaces, with customization happening in the ensuing force feedback step and with an eventual syntax depending on the particular interface's constraints and resources. Also, by providing a semantic base built upon gestures taken both from music and from everyday life, we expect to provide the necessary freedom that any expressive interface should have.

2.3.2 Design and Deployment of the Haptic Feedback

Our proposition is that the design of the haptic feedback should be governed by a perceptual framework, separating each haptic channel according to the use and meaning it possesses in the real-life context. Research in this most substantial phase will focus on assigning an appropriate haptic environment (consisting of an active

tactile or force behavior) to any useful perceptual (auditory) music parameters like pitch, timbre, loudness, rhythm or tempo, just to mention some. Only when these responses and environments are tuned and tested through user studies, will they be matched to the gestures defined with the semantic base as the ones most appropriate for controlling those parameters. Thus, the haptic environment will tend to be associated with the identity and value of the parameter, and the gesture with the action to be performed upon it.

For instance, if we find that the control action of increasing pitch is best controlled by presenting a force that opposes the user's movements (providing a resistance that enables more precise control over pitch), then first the designer would tune the force feedback parameters to a point that they feel comfortable. Next, he/she would match this feedback with the gesture for controlling the pitch, and further adjust it until this combination feels right. The haptic feedback design deals only with the best way to perceive a particular music parameter through a haptic channel, while carrying out the semblance of a particular gesture.

The design of each flavor of haptic feedback is done in separate stages and is incorporated into the general process at different times. We base this approach on the possibility of separating the force feedback from the tactile feedback channels to address different interface features: the former to foster controllability and the latter to foster the feeling of using a music instrument.

We showed how in traditional (acoustic) music instruments the vibrations felt by the performer depend on the generated sound. Consequently, a major percentage of the vibrotactile feedback in our case will be extracted from the sound being generated by the gestures performed over the instrument. We would like to experiment the consequences of providing cues to several music parameters like rhythm or tempo on top of that "generic" vibrotactile feedback.

A set of user studies will help the final fine-tuning of the haptic feedback parameters (still in isolation from one another, but in conjunction with the result obtained from previous stages like gestures associated with those parameters / actions). They will serve to evaluate ensuing performance in parameter control, as well as to identify relevant perceptual limits. With these user studies we plan to gather a list of some representative restrictions/suggestions pertaining to an appropriate division of control/perception functionality between the two haptic channels.

2.3.3 Merging of Tactile/Force Feedback

There are two mixing stages in the proposed research, and each one fulfills one particular objective.

The first mixing stage has the objective of enhancing the controllability of a gesture interface by providing some impedance (force feedback) to the gestures controlling the music parameters. This stage accepts two inputs: a general set of gestures designed to control the music parameters defining the music/expression space, and a set of forces perceptually fine-tuned to achieve the best apprehension (identity and setting) of the same set of music parameters. In this first mixing stage, the main objective is to create a force space that matches the music/expression space in a way that each music parameter could be controlled effectively. As the gestures included in the semantic base are not grounded, we anticipate some changes but not enough to require us to iterate on the basic gestures themselves.

The second mixing stage is aimed at providing the feeling of playing a music instrument in an interface already tuned for control.

This stage presents a higher risk, since some perceptual overloading might occur. However, there may be ways to reduce overloading by using masking and or neglecting a haptic cue that provides less information. The goal here is not to optimize the design for everything that could be controlled, but to achieve the best ratio of controllability and perception of a music instrument. Just as some music instruments are more expressive and versatile than others, the same situation occurs here. We are aiming to obtain the best performance features with the resources at hand.

3 Conclusions

We have proposed and elaborated on a methodology aimed to address the potential benefits of and challenges inherent in including haptic feedback in expressive control interfaces, with a case-study focus on computer music performance. Through a perceptual analysis of the musician-instrument interaction, a more informed design could be arranged whereby feedback signals conveyed to the performer could be assigned to the perception channels (vibrotactile or force feedback) that represent them in traditional contexts.

This methodology is being used at the SPIN Research Group (The University of British Columbia).

Acknowledgments. The main author would like to acknowledge the feedback received from Drs. Keith Hamel, Alan Kingstone and Michael van de Panne as well as UBC's SPIN Research Group, while developing this methodology.

References

1. Askenfelt, A., Jansson, E.V.: On vibration sensation and finger touch in stringed instrument playing. *Music Perception* 9(3), 311–350 (1992)
2. Bongers, B.: The use of active tactile and force feedback in timbre controlling electronic instruments. In: *Proceedings of ICMC*, pp. 171–174 (1994)
3. Brouwer, I., MacLean, K.E., Hodgson, A.: Simulating Cheap Hardware: A Platform for Evaluating Cost-Performance Trade-Offs in Haptic Hardware Design. In: *Proc. of IEEE Robotics and Automation (ICRA 2004)*, pp. 770–775 (2004)
4. Cadoz, C.: Le geste canal de communication homme/machine. *La communication "instrumentale. Technique et science informatique* 13(1), 31–61 (1994)
5. Chafe, C.: Tactile Audio Feedback. In: *Proceedings of ICMC, Tokyo*, (September 1993)
6. Chu, L.: Haptic Feedback in Computer Music Performance. In: *Proceedings of ICMC*, pp. 57–58 (1996)
7. Cook, P.: Hearing, feeling and playing: masking studies with trombone players. In: *Proceedings of the 4th International Conference on Music Perception and Cognition*, pp. 513–518. McGill University (1996)
8. Essl, G., O'Modhrain, S.: Scrubber: an interface for friction-induced sounds. In: *Proceedings of NIME 2005, Vancouver, Canada*, pp. 70–75 (2005)

9. Finney, S.A.: Auditory feedback and musical keyboard performance. *Music Perception* 15(2), 153–174 (Winter 1997)
10. Gillespie, B.: The Touchback Keyboard. In: Proceedings of ICMC, October. 14–18, pp. 447–448 (1992)
11. Gillespie, B.: Haptic display of systems with changing kinematic constraints: The Virtual Piano Action. Ph.D. dissertation, Stanford University, <http://www-personal.umich.edu/~brentg/Publications/Thesis/thesis.html>
12. Lederman, S.J., Klatzky, R.L.: Haptic aspects of motor control. In: Boller, F., Grafman, J. (eds.) *Handbook of neuropsychology*, Vol. pp. 131–148. Elsevier, New York (1997)
13. Leman, M., Styns, F.: Sound, sense, and music mediation: A historical/philosophical perspective. In: Marc Leman, Damien Cirrotteau, eds. *Sound to Sense, Sense to Sound: A State-of-the-Art. S2S² Project Deliverable. Version 0.10 (CVS:Nov.9 (2005))*, http://www.s2s2.org/docman/task,doc_download/gid,70/Itemid,65
14. O’Modhrain, M.S., Chafe, C.: Incorporating Haptic Feedback into Interfaces for Music Applications. In: 8th International Symposium on Robotics with Applications, ISORA 2000, World Automation Congress WAC (2000)
15. O’Modhrain, S., Essl, G.: PebbleBox and CrumbleBag: tactile interfaces for granular synthesis. In: Proceedings of NIME 2004, Hamamatsu, Shizuoka, Japan, June 03–05, pp. 74–79 (2004)
16. Repp, B.H.: Effects of auditory feedback deprivation on expressive piano performance. *Music Perception* 16(4), 409–438 (Summer, 1999)
17. Rován, J., Hayward, V.: Typology of Tactile Sounds and their Synthesis in Gesture-Driven Computer Music Performance. In: Wanderley, M., Battier, M. (eds.) *Trends in Gestural Control of Music*. Editions IRCAM, Paris (2000)

Real-Time Gesture Recognition, Evaluation and Feed-Forward Correction of a Multimodal Tai-Chi Platform

Otniel Portillo-Rodriguez, Oscar O. Sandoval-Gonzalez, Emanuele Ruffaldi,
Rosario Leonardi, Carlo Alberto Avizzano, and Massimo Bergamasco

PERCRO, Perceptual Robotics Laboratory,
Scuola Superiore Sant'Anna, Pisa Italy

Abstract. This paper presents a multimodal system capable to understand and correct in real-time the movements of Tai-Chi students through the integration of audio-visual-tactile technologies. This platform acts like a virtual teacher that transfers the knowledge of five Tai-Chi movements using feed-back stimuli to compensate the errors committed by a user during the performance of the gesture. The fundamental components of this multimodal interface are the gesture recognition system (using k-means clustering, Probabilistic Neural Networks (PNN) and Finite State Machines (FSM)) and the real-time descriptor of motion which is used to compute and qualify the actual movements performed by the student respect to the movements performed by the master, obtaining several feedbacks and compensating this movement in real-time varying audio-visual-tactile parameters of different devices. The experiments of this multimodal platform have confirmed that the quality of the movements performed by the students is improved significantly.

Keywords: Multimodal Interfaces, real-time 3D time-independent gesture recognition, real-time descriptor, vibrotactile feedback, audio-position feedback, Virtual Reality and Skills transfer.

1 Introduction

The learning process is one of the most important qualities of the human being. This quality gives us the capacity to memorize different kind of information and behaviors that help us to analyze and survive in our environment. Approaches to model learning have interested researches since long time, resulting in such a way in a considerable number of underlying representative theories.

One possible classification of learning distinguishes two major areas: Non-associative learning like habituation and sensitization, and the associative learning like the operant conditioning (reinforcement, punish and extinction), classical conditioning (Pavlov Experiment), the observational learning or imitation (based on the repetition of a observed process) [1], play (the perfect way where a human being can practice and improve different situations and actions in a secure environment) [2], and the multimodal learning (dual coding theory) [3].

Undoubtedly, the imitation process has demonstrated a natural instinct action for the acquisition of knowledge that follows the learning process mentioned before. One

example of multimodal interfaces using learning by imitation in Tai-chi has been applied by the Carnegie Mellon University in a Tai-Chi trainer platform [4], demonstrating how through the use of technology and imitation the learning process is accelerated.

The human being has a natural parallel multimodal communication and interaction perceived by our senses like vision, hearing, touch, smell and taste. For this reason, the concept of Human-Machine Interaction HMI is important because the capabilities of the human users can be extended and the process of learning through the integration of different senses is accelerated [5] [6] [7]. Normally, any system that pretends to have a normal interaction must be as natural as possible [8] [9]. However, one of the biggest problems in the HMI is to reach the transparency during the Human-Machine technology integration.

In such a way, the multimodal interface should present information that answers to the “why, when and how” expectations of the user. For natural reasons exists a remarkable preference for the human to interact multimodally rather than unimodally. This preference is acquired depending of the degree of flexibility, expressiveness and control that the user feels when these multimodal platforms are performed [9]. Normally, like in real life, a user can obtain diverse information observing the environment. Therefore, the Virtual Reality environment (VR) concept should be applied in order to carry out a good Human-Machine Interaction. Moreover, the motor learning skills of a person is improved when diverse visual feedback information and correction is applied [10].

For instance the tactile sensation, produced on the skin, is sensitive to many qualities of touch. Lieberman and Breazeal [11] carried out, for first time, an experiment in real time with a vibrotactile feedback to compensate the movements and accelerate the human motion learning. The results demonstrate how the tactile feedback induces a very significant change in the performance of the user. In the same line of research Boolmfield performed a Virtual Training via Vibrotactile Arrays[12].

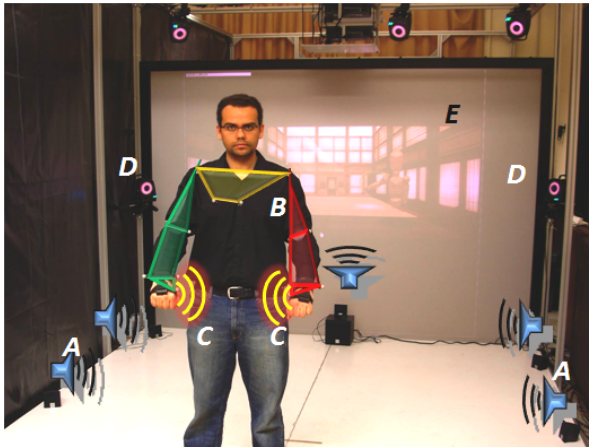


Fig. 1. Multimodal Platform set up, A) 3D sound, B) Kinematics Body C) Vibrotactile device (SHAKE) D) Vicon System E) Virtual Environment

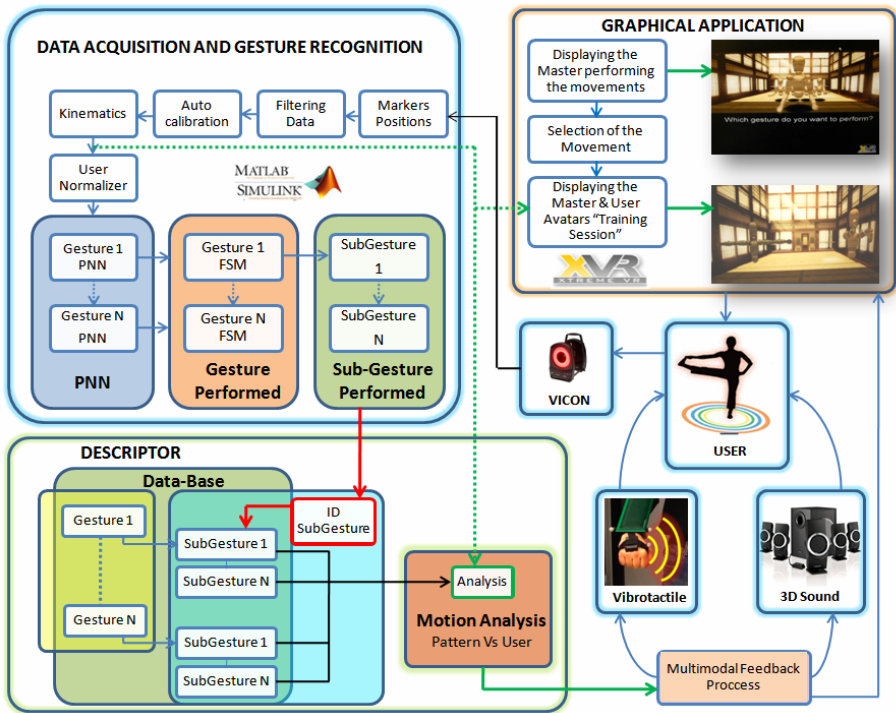


Fig. 2. Architecture of the Multimodal Tai Chi Platform System

Another important perception variable is the sound because this variable can extend the human perception in Virtual Environments. The modification of parameters like shape, tone and volume in the sound perceived by the human ear [13], is a good approach in the generation of the description and feedback information in the human motion.

Although a great grade of transparency and perception capabilities are transmitted in a multimodal platform, the intelligence of the system is, unquestionably, one of the key parts in the Human-Machine interaction and the transfer of a skill. Because of the integration, recognition and classification in real-time of diverse technologies are not easy tasks, a robust gesture recognition system is necessary in order to obtain a system capable to understand and classify what a user is doing and pretending to do.

2 System Implementation

This paper presents a multimodal interface that teaches to novel students, five basic tai chi movements. Each movement is indentified and analyzed in real time by the gesture recognition system. The gestures performed by the users are subdivided in n-states (time-independent) and evaluated step-by-step in real time by the descriptor system. Finally, the descriptor executes audio-visual-tactile feed-back stimuli in order to correct the user’s movements. Fig. 1 presents the interface that is composed

by: The hardware and software of the 3D tracking optical system (VICON), the gesture recognition system and the description of motion (both running in Matlab Simulink), a graphical scenario developed in XVR, a 3D sound system and the wireless vibrotactile devices (SHAKE). The general architecture of the multimodal platform is shown in Fig. 2.

2.1 Data Acquisition

The motion of the Tai-Chi student was tracked with the VICON system. This system is an optical device which provides millimeter accuracy in the 3D space through the use of passive reflective markers attached to the body at 300Hz of sampling frequency. Sometimes, due to the markers obstructions in the human motion, the data information is lost. For this reason, the “cleaning algorithm” described in [14], was implemented. An inverse kinematics of fourteenth DOFs represented by the upper part of the body is computed. A calibration process is completely required in order to identify the actual position of the markers and adjust the kinematics model to the new values. Therefore, a fast (1ms) autocalibration process was designed in order to obtain the initial position of the markers of a person placed in a military position called “stand at attention”. The algorithm checks the dimension of his/her arms and the position of the markers. The angles are computed and finally this information is compared with to the ideal values in order to compensate and normalize the whole system.

2.2 Real-Time Gesture Recognition Process

In order to recognize the gesture performed by the user, a state space model approach was selected [15][16]. Normally, the principal problem to model a gesture in the state based approach, is the characterization of the optimal number of states and the establishment of their boundaries. For each gesture, the training data is obtained concatenating the data of five demonstrations. A dynamic k -means clustering on the training data defines the number of states and their spatial parameters of the gesture without temporal information [17]. This information from the segmented data is then added to the states and finally the spatial information is updated. This produces the state sequence that represents the gesture. The analysis and recognition of this sequence is performed using a simple Finite State Machine (FSM) [18], instead of use complex transitions conditions which depend only of the correct sequence of states for the gesture to be recognized and eventually of time restrictions i.e., minimum and maximum time permitted in a given state.

The novel idea is to use for each gesture a PNN to evaluate which is the nearest state (centroid in the configuration state) to the current input vector that represents the user’s body position. The input layer has the same number of neurons as the input vector and the second layer has the same quantity of hidden neurons as states have the gesture. In our architecture (Fig. 3), each class node is connected just to one hidden neuron and the number of states (where the gesture is described) defines the quantity of class nodes. Finally, in the last layer, the class (state) with the highest summed

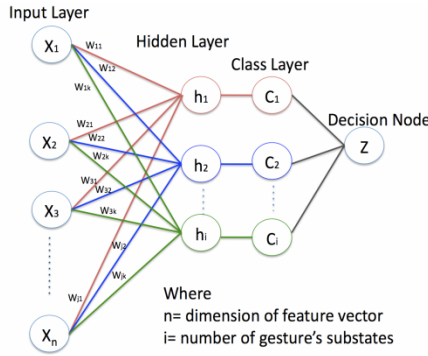


Fig. 3. PNN architecture used to estimate the most similar gesture's state from the current user's body position

activation is computed. A number of 12 variables were used in our configuration space: There are 2 distances between hands and 2 between elbows. 2 Vectors created from the XYZ position from the hands to the chest and 2 Vectors created from the XYZ positions from the elbows to the chest.

2.3 Real-Time Descriptor Process

The comparison and qualification in real-time of the movements performed by the user is computed by the descriptor system. In other words, the descriptor analyzes the differences between the movements executed by the expert and the movement executed by the student, obtaining the error values and generating the feedback stimuli to correct the movement of the user. Each pattern movement is characterized for a sequence of states which is formed by 18 variables and performs the comparison of the following information: *12 Angles:* Elbows(2), Wrists(4) and Shoulders(6), *2 Distances:* Distance between hands(1) and elbows(1) and *4 Positional vectors:* 2 vectors created from the XYZ position of the hands to the chest and 2 Vectors created from the XYZ positions of the elbows to the chest. Each state or subgesture is recognized in real time by the gesture recognition system during the performance of the movement. Using the classic feed-back control loop during the experiments was observed that the user feels a delay in the corrections. For that reason, a feed-forward strategy was selected to compensate this perception. In this methodology when a user arrives at one state of the gesture, the descriptor system creates n-substates and carries out an interpolation process to compare the actual values with respect to the values in the sub-state ($n+1$) of the pattern value, creating a feed-forward loop which estimates in advance the next correction values of the movement. The error is computed by:

$$\theta_{error} = [P(n + 1) - U(n)] * F_n \tag{1}$$

Where θ_{error} is the difference between the pattern and the user, P is the pattern value, U is the user value, F_u is the normalize factor and n is the actual state.

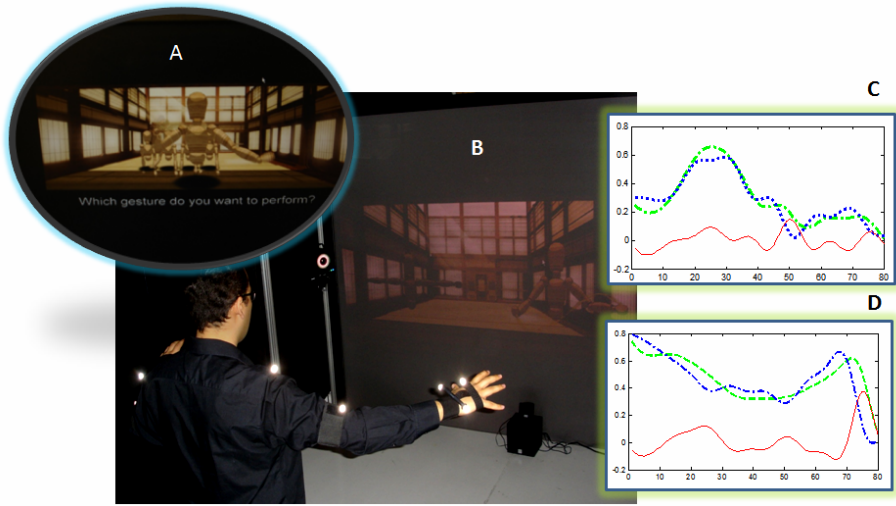


Fig. 4. VR environment , A) Initial Screen, 5 avatars performing Tai-Chi movements, B) Training session, two avatars, one is the master and second is the user. C) Distance of the Hands, D) Right Hand Position.

2.4 Virtual Reality Platform

The virtual environment platform which provides the visual information to the user was programmed in XVR. There are 3 different sequences involved in this scenery. The first one is the initial screen that shows 5 avatars executing different Tai Chi movements. When a user tries to imitate one movement, the system recognizes the movement through the gesture recognition algorithm and passes the control to the second stage called “training session”. In this part, the system visualizes 2 avatars, one represents the master and the other one is the user. Because learning strategy is based on the imitation process, the master performs the movement one step forward to the user. The teacher avatar remains in the state($n+1$) until the user has reached or performed the actual state(n). With this strategy the master gives the future movement to the user and the user tries to reach him. Moreover, the graphics displays a virtual energy line between the hands of the user. The intensity of this line is changing proportionally depending on the error produced by the distance between the hands of the student. When a certain number of repetitions has been performed, the system finishes the training stage and displays a replay section which shows all the movements performed by the student and the statistical information of the movement’s performance. Fig. 4 (A)(B) shows the virtual Tai-Chi environment.

2.5 Vibrotactile Feedback System

The SHAKE device was used to obtain wireless feedback vibrotactile stimulation. This device contains a small motor that produces vibrations at different frequencies. In this process, the descriptor obtains the information of the distance between the hands, after this, the data is compared with the pattern and finally sends a proportional

value of the error. The SHAKE varies proportionally the intensity of the vibration according to error value produced by the descriptor (1 Hz – 500 Hz). This constraint feedback is easy to understand for the users when the arms have reached a bad position and need to be corrected. Fig. 4 (C) shows the ideal distance between the hands (green), the distance between the hands performed by the user (blue) and the feedback correction (red).

2.6 Audio Feedback System

The position of the arms in the X-Y plane is analyzed by the descriptor and the difference in position between the pattern and the actual movement in each state of the movement is computed. A commercial Creative SBS 5.1 audio system was used to render the sound through 5 speakers (2 Left, 2 Right, 1 Frontal) and 1 Subwoofer. In this platform was selected a background soft-repetitive sound with a certain level of volume. The sound strategy performs two major actions (volume and pitch) when the position of the hands exceeds the position of the pattern in one or both axes. The first one increases, proportionally to the error, the volume of the speakers in the corresponding axis-side (Left-Center-Right) where is found the deviation and decreases the volume proportionally in the rest of the speakers. The second strategy varies proportionally the pitch of the sound (100-10KHz) in the corresponding axis-side where was found the deviation. Finally, the user through the pitch and the volume can obtain information which indicates where is located the error and its intensity in the space.

3 Experimental Results

The experiments were performed capturing the movements of 5 Tai-Chi gestures (Fig. 5) from 5 different subjects. The tests were divided in 5 sections where the users performed 10 repetitions of the each one of the 5 movement performed. In the first section was avoided the use of technology and the users performs the movement in a traditional way, only observing a video of a professor performing one simple tai-chi movement. The total average error TAVG is calculated in the following way:

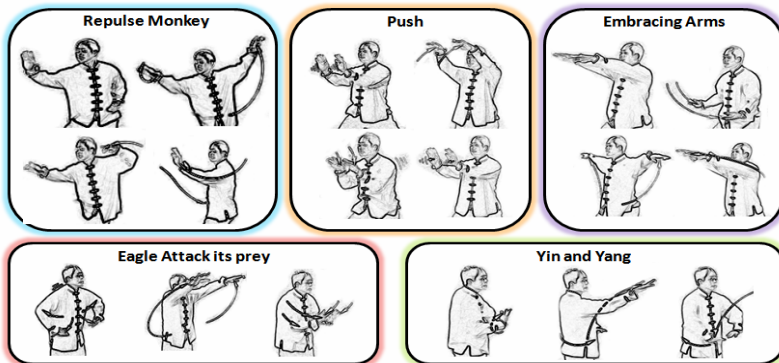


Fig. 5. The 5 Tai-Chi Movements

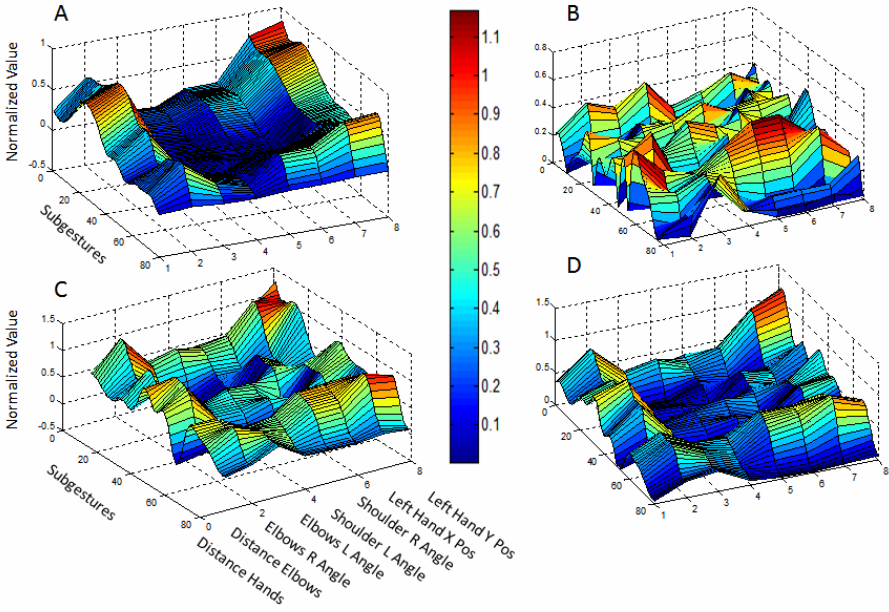


Fig. 6. Variables of Gesture 1, A) Pattern Movement, B) Movement without feedback, C) Movement with Visual feedback and D) Signals with Audio-Visual-Tactile feedback.

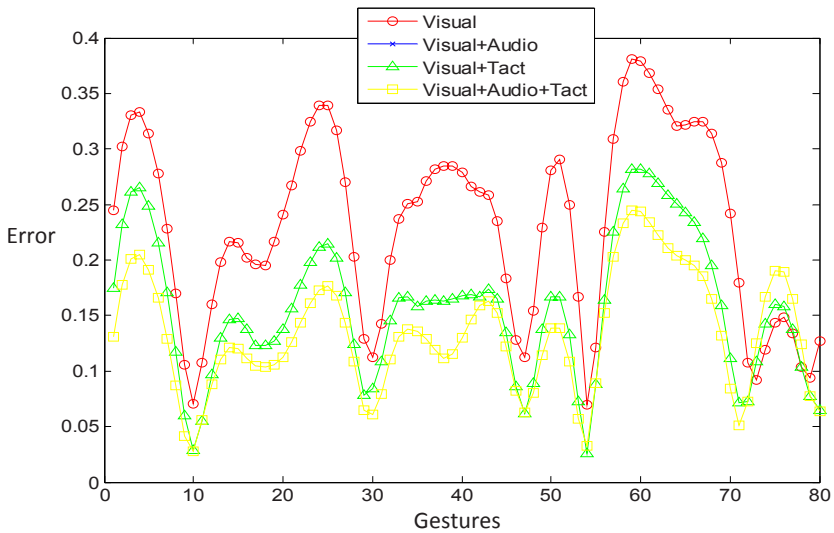


Fig. 7. Average Errors

$$TAVG = \frac{1}{Ns} \sum_{s=0}^{Ns} \frac{1}{n} \sum_{i=0}^n (\theta_{Teacher} - \theta_{Student}) \quad (2)$$

Where Ns is the total number of subjects, n is the total number of states in the gesture and θ is the error between the teacher movement and the student.

Fig. 6 (A) shows the ideal movements (Master Movements) of the gesture number 1 and (B) represents the TAVG of the gesture 1 executed by the 5 subjects without feedback. The TAVG value the 5 subjects without feedback was around 34.79% respect to the ideal movement.

In the second stage of the experiments, the Virtual Reality Environment was activated. The TAVG value for the average of the 5 subjects in the visual feedback system presented in Fig. 6(C) was around 25.31%. In the third section the Visual-Tactile system was activated and the TAVG value was around 15.49% respect to the ideal gesture. In the next stage of the experiments, the visual- 3D audio system was performed and the TAVG value for the 5 subjects in the audio-visual feedback system was around 18.42% respect to the ideal gesture. The final stage consists in the integration of the audio, vibrotactile and visual systems. The total mean error value for the average of the 5 subjects in the audio-visual-tactile feedback system was around 13.89% respect to the ideal gesture. Fig. 6 (D) shows the results using the whole integration of the technologies. Finally, Fig. 7 presents an interesting graph where the results of the four experiments are indicated. In one hand, as it was expected, the visual feedback presented the major error. In the other hand the integration of audio-visual-vibrotactile feedback has produced a significant reduction of the error of the users.

4 Conclusions

A novel methodology of a real-time gesture recognition and descriptor used in a multimodal platform with audio-visual-tactile feedback system was presented in this paper. The aim to obtain a robust gesture recognition system capable to recognize 5 complex gestures and divide them in different subgestures was fulfilled. Moreover, the function of the real-time descriptor offers the possibility to analyze and evaluate, in a separate and integrate way, the behavior of movements from the different variables related to the feed-back system (audio, vision and tact). The results of the experiments have shown that although the process of learning by imitation is really important, there is a remarkable improvement when the users perform the movements using the combination of diverse multimodal feedbacks systems.

5 Future Work

Once the multimodal platform has demonstrated the feasibility to perform the experiments related to the transfer of a skill in real-time, the next step will be focused in the implementation of a skill methodology which consists, in a brief description, into acquire the data from different experts, analyze their styles and the descriptions of the most relevant data performed in the movement and, through this information, select a certain lessons and exercises which can help the user to improve his/her movements.

Finally it will be monitored these strategies in order to measure the progress of the user and evaluate the training. These information and strategies will help us to understand in detail the final effects and repercussions that produces each multimodal variable in the process of learning.

References

1. Byrne, R., Russon, A.: Learning by imitation: a Hierarchical Approach. *Behavioral and Brain Sciences* 21, 667–721 (1998)
2. Spitzer, M.: *The mind within the net: models of learning, thinking and acting*. The MIT press, Cambridge (1998)
3. Viatt, S., Kuhn, K.: Integration and synchronization of input modes during multimodal human-computer interaction. In: *Proc. Conf. Human Factors in Computing Systems CHI* (1997)
4. Tan Chua, P.: Training for physical Tasks in Virtual Environments: Tai Chi. In: *Proceedings of the IEEE Virtual Reality* (2003)
5. Cole, R., Mariani, J.: *Multimodality*. In: *Survey of the State of the Art of Human Language Technology*, Carnegie Mellon University, Pittsburgh, PA (1995)
6. Sharma, R., Pavlovic, V., Huang, T.: Toward Multimodal Human-Computer Interface. *Proceedings of the IEEE* 86(5), 853–869 (1998)
7. Akay, M., Marsic, I., Medl, A.: A System for Medical Consultation and Education Using Multimodal Human/Machine Communication. *IEEE Transactions on information technology in Biomedicine* 2 (1998)
8. Hauptmann, A.G., McAvinney, P.: Gesture with Speech for Graphics Manipulation. *Man-Machines Studies* 38 (1993)
9. Oviatt, S.: User-Centered Modeling and Evaluation of Multimodal Interfaces. *Proceedings of the IEEE* 91 (1993)
10. Bizzi, E, Mussa-Ivaldi, F.A. and Shadmehr, R. : System for human trajectory learning in virtual environments. US Patent No. 5,554,033 (1996)
11. Lieberman, J., Breazeal, C.: Development of a wearable Vibrotactile FeedBack Suit for Accelerated Human Motor Learning. In: *IEEE International Conference on Robotics and Automation* (2007)
12. Bloomfield, A., Badler, N.: Virtual Training via vibrotactile arrays. *Teleoperator and Virtual Environments* 17 (2008)
13. Hollander, A.J., Furness III, T.A.: Perception of Virtual Auditory Shapes. In: *Proceedings of the International Conference on Auditory Displays* (1994)
14. Qian, G.: A gesture-Driven Multimodal Interactive Dance System. In: *IEEE International Conference on Multimedia and Expo ICME* (2004)
15. Bobick, W.: State-Based Approach to the Representation and Recognition of Gesture. *Pattern Analysis and Machine Intelligence, IEEE Transactions* 19 (1997)
16. Farmer, J.: State-space reconstruction in the presence of noise. *Physics* (1991)
17. Jain, A.K., Murty, M.N., Flynn, P.J.: Data Clustering: A review. *ACM Computing Surveys* 31 (1999)
18. Hong, P., Turk, M.: Gesture Modeling and Recognition Using Finite State Machines. In: *Proceedings of the Fourth IEEE International Conference on Automatic Face and recognition* (2000)

A System for Multimodal Exploration of Social Spaces

Victor V. Kryssanov^{1,*}, Shizuka Kumokawa², Igor Goncharenko³,
and Hitoshi Ogawa¹

¹ College of Information Science and Engineering, Ritsumeikan University
kvvictor@is.ritsumei.ac.jp

² Graduate School of Science and Engineering, Ritsumeikan University,
1-1-1, Noji-Higashi, Kusatsu, Shiga 525-8577, Japan

³ 3D Incorporated, Urban Square Yokohama Bldg. 2F 1-1 Sakae-cho, Kanagawa-ku,
Yokohama, Kanagawa, Japan

Abstract. This paper describes a system developed to help people explore local communities by providing navigation services in social spaces created by members of the communities. Just as a community's social space is formed by communication and knowledge-sharing practices, the proposed system utilizes data of the corresponding social network to reconstruct the social space, which is otherwise not physically perceptible but imaginary and yet experiential and learnable. The social space is modeled with an agent network, where each agent stands for a member of the community and has knowledge about expertise and personal characteristics of other members. An agent can gather information, using its social "connections," to find community members most suitable to communicate to in a specific situation defined by the system's user. The system then deploys its multimodal interface, which operates with 3D graphics and haptic virtual environments and "maps" the social space onto a representation of the relevant physical space, to advise the user on an efficient communication strategy for the given community. A prototype of the system is built and used in a pilot study. The study results are briefly discussed, conclusions are drawn, and implications for future work are formulated.

Keywords: Social navigation, agent network, multimodal interface.

1 Motivation

Since the advent of computer age several decades ago, the role of various information systems in human knowledge sharing and proliferation has been accelerating. At the same time, however, the bulk of information learned by people in their lifetimes still never appears in a database or on the Internet but is readily available to members of various local communities, such as families, school students and alumni, indigenous people, company employee, and the like. This information is typically conveyed via word-of-mouth in conversations on an individual, person-to-person basis. While the modern information technologies traditionally focus on asynchronous mass-communication and deliver a vast array of tools (e.g. electronic libraries and search engines) supporting this form of information exchange, little has been done to assist the essentially personified

* Corresponding author.

and synchronous communication occurring daily, as we quire a teacher at a school, ask a local for directions, or seek advice from a friend or the “best expert” (e.g. a doctor or lawyer) in a field. Even though existing computer systems do provide for person-to-person information exchange, their support does not go far beyond, say, a postal service that allows people already socially connected to communicate. Whether we walk on a street or chat using an instant messenger, or else write to a forum of a social network system, our chances of obtaining information of interest are roughly the same. It is our abilities to navigate in social spaces, which are, at best, partly known, and to initiate and maintain communication at a level of synchronicity optimal for given time constraints that determine the success or otherwise of an information quest. None of the present-day information systems and “e-services” known to the authors targets supporting this essentially “interhuman” navigation process. Besides, the very concepts of social space and communication synchronicity, although not totally alien in computer sciences, are presently discussed as quite theoretical and speculative rather than as something that would strongly affect and be practically used in information system design and development [1,4].

The presented study aims at the creation of an information service to allow people to navigate in (unknown) social environments and help locate “carriers” of specific information (i.e. advisers) that would be approached in a particular situation. This paper describes a multi-agent information system “SoNa” (*Social Navigator*) developed to assist the users in exploring a social space formed by a local community and in obtaining information of interest from members of the community.

In line with the most common understanding of the social space concept (see [6,7]), the proposed system reproduces in a 3D virtual reality (a relevant fragment of) the physical space together with members of the local community present in the space at the moment. Unlike the physical proximity, social relationships (e.g. “social distance” or “friendship”) are usually not directly perceived in real life, but are inferred and “felt” from (collective and individual) communicative experiences. A haptic environment including a force display is then used to convey important parts of the community’s communication practices – the “social knowledge” – to the user via the “subconscious” tactile communication channel. An agent network is created and used by the system to deal with the social knowledge. This network represents a real social network of the community, and the agents exchange information by communicating with their “socially connected” counterparts in the same way as people do it in the real world. Each agent in the network has parameters indicating whether the corresponding member is sociable, friendly, can be trusted, can afford to communicate (e.g. in terms of time) and is currently reachable (e.g. physically or via e-mail). Apart from exploration of the social space in various modalities and under different contexts, the user can use the system as a navigator in his or her search of a community member who would be approached with a specific information request.

In the next section, the design of the proposed system is presented. Section 3 then describes a working prototype of the system implemented in the study and elaborates on the multimodal user interface and user-system interaction. Section 4 gives a short account of a pilot study of applying the developed prototype in practice. Finally, Section 5 concludes the paper.

2 System Design

Navigation in an environment, whether physical or virtual, can generally be defined as a four-stage iterative process [9]: 1) perception of the environment, 2) reconciliation of the perception and cognition, 3) deciding on whether the goal has been reached, and 4) selecting the next action. Among these stages, only the first two directly depend on information about the environment and can thus be supported with an information system [5]. To provide for navigation in the social space, the proposed system has five functions: 1) creating profiles of individuals embedded in the social space and constructing a network of agents that reflects the community’s social network, 2) receiving the user’s request and gathering the agents’ knowledge, 3) extracting information which meets the user’s needs, 4) displaying the social space in 3D graphics and haptic virtual environments, and 5) updating the states of the agents.

Fig. 1 depicts the architecture of the system. The agent manager creates the agent network using data stored in the database. The multimodal interface provides for the interaction of the user with other parts of the system and delivers information for the navigation process. When the recommender system receives a request through the multimodal interface, it selects an appropriate agent and sends a query to this agent to gather information in the network. For information gathering, an efficient communication algorithm is implemented (see [10] for details of the algorithm). Once the recommender system receives responses from all agents in the network, it analyzes the obtained information and sends results of the analysis to the multimodal interface. The multimodal interface presents the information sought as well as all the relevant segments of the social and physical spaces.

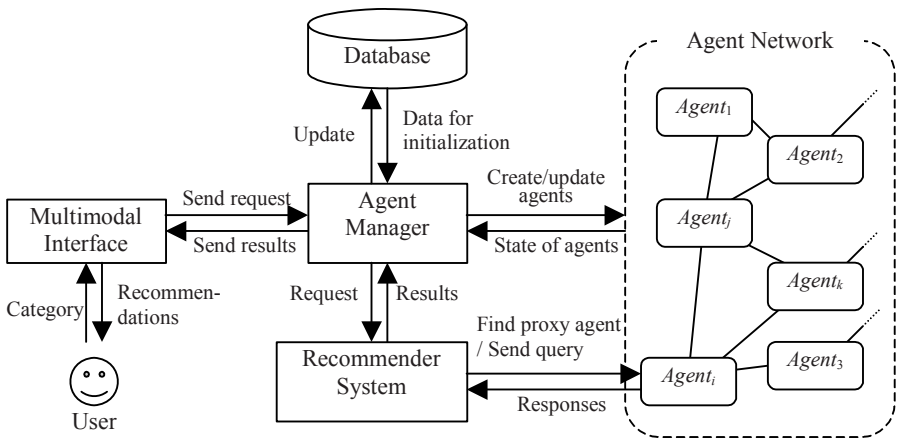


Fig. 1. System architecture

2.1 The Agent Network

The agent network is composed of the same number of agents as the number of members in the community in focus, and the agents are connected to their “acquaintance

agents” just as the corresponding people are socially connected in the real world. To assemble the agents into a network, the following information is required: profiles with individual characteristics of the community members, personal network data of the community members, and each member’s communicative experiences in respect to some pre-defined “objects” which are other members.

First, the agent manager creates N_A agents, denoted a_1, a_2, \dots, a_{N_A} ; N_A is the number of members in the community. Each agent’s profile, which is registered in the database, initially includes only static data, such as member’s name, gender, and “permanent” location/address in the physical space. This static data is set as the agent’s parameters. Next, the agent manager informs agents (by attaching relevant descriptions – keywords, etc.) about the expertise of the corresponding members in the real world. Members with expertise are called objects and denoted o_1, o_2, \dots, o_{N_o} , N_o is the number of the objects. The objects are thus members who have been evaluated by members of the community in respect to a specific category of knowledge c_k . The evaluation is performed by rating the objects as relevant or irrelevant when information sought falls into the category c_k . Totally, there are N_c pre-defined categories denoted c_1, c_2, \dots, c_{N_c} . The rate of O_j is represented as r_j . In the current implementation of the agent model, we assume the binary rate: $r_j = 1$ for the positive evaluation, and $r_j = -1$ otherwise.

Each agent is connected to its “acquaintance agents,” using the members’ personal network data. This data has the structure of an undirected graph, and the state of linkage between the nodes (i.e. agents) is represented, using an adjacency matrix. The matrix is constructed via mutual certification of pairs of agents. Each agent then receives two dynamic attributes: agents-acquaintances called “neighbors” and “trust values” for the neighbors. A trust value between agents a_i and a_j is expressed as T_{a_i, a_j} , which is a real number fluctuating between 0 (no information / low trust) and 1 (full trust). The dynamics of T_{a_i, a_j} is specified with the following equations ($t = 0$ corresponds to the moment when the agent network is initialized):

$$\begin{aligned} \tilde{T}_{a_i, a_j}(t=0) &= 0, \\ \tilde{T}_{a_i, a_j}(t+1) &= \begin{cases} \gamma \tilde{T}_{a_i, a_j}(t) + (1-\gamma)r_k, & r_k > 0; \\ (1-\gamma)\tilde{T}_{a_i, a_j}(t) + \gamma r_k, & r_k < 0; \end{cases} \quad (1) \\ T_{a_i, a_j}(t+1) &= \frac{1 + \tilde{T}_{a_i, a_j}(t+1)}{2}, \quad t = 0, 1, \dots \end{aligned}$$

r_k is a rate shared by (i.e. common for) the two members, and parameter γ determines to what extent the previous trust value affects the new trust value. When γ is greater than 0.5, the trust value increases slowly, and decreases quickly that is the “trust dynamics” often observed in real social networks [10].

The trust value between two agents that are not directly connected is calculated as the product of all of the trust values in the path connecting the two agents. At any time, members can add new rates for objects into the knowledge of their respective agents. The calculated trust values are used by the recommender system when there

are more than one person to recommend under the same conditions. The data of the recommended object is sent to the user interface with a weight w calculated as follows:

$$w = \frac{\exp(\beta \hat{T}_{ai, aj})}{\sum_R \exp(\beta \hat{T}_{ai, ak})}, \quad (2)$$

$$\hat{T}_{ai, aj} = \frac{1}{2} \ln \left(\frac{1 + 2(T_{ai, aj} - 0.5)}{1 - 2(T_{ai, aj} - 0.5)} \right). \quad (3)$$

In formula (2), summation is done over R , the set of all responses received from agents $a_k, k = 1, \dots, N_R, N_R$ is the number of responses, to the specific request; parameter β determines to what degree the weight accounts for the trust value: when β is 0, all weights are the same, and when β is close to 1, an object rated by an agent with a higher trust value has a greater weight.

2.2 Agent Functions and the Recommender System

Whenever the system user is not a member of the community, the recommender system accesses the agent network to find an agent with a profile most similar to the user's self-description, and makes this agent the user's proxy. The user, whose (proxy) agent is a_i , inputs a category of her/his enquiry, c_j . The recommender system sends the user's query to a_i in the agent network, where it is relayed by a_i to its neighbors as $query(a_i, c_j)$. When a neighbor of the agent receives the query, it checks if it has knowledge about objects rated in the category c_j . If the neighbor agent finds a rated object, it sends a response to the agent a_i . The response is formed as $response(a_i, a_k, c_j, (o_l, r_l), T_{ai, ak})$, a_k stands for the agent, which sent the response, and o_l is the object, which is rated r_l by a_k in category c_j ; $T_{ai, ak}$ is the trust value between a_i and a_k . If the neighbor does not have knowledge about objects in the given category, it further transmits the query to its neighbors. These latter agents process the query in the same way as described above. To prevent unnecessary communications, every agent, which has once processed a query, ignores this query when it is received repeatedly. Usually, there are many paths between agents a_i and a_k in the agent network, but the generated responses always pass through the same path as the query does.

When the information-gathering algorithm terminates, the agent, which originally sent the query, has a set of responses from the community members. The next step is selecting objects to recommend, which meet the user's criteria, from the available set. In the set, there are sometimes objects impossible to put in use at the given moment, regardless of how strongly the members would recommend them. For example, when the user needs to meet an adviser immediately, if the system recommends a person who is currently not reachable or who can only speak a foreign language, such a recommendation would have little practical value. The system filters the responses to remove any potentially useless recommendations and, by doing so, tries to balance the synchronicity of the expected communication.

3 Developed Prototype

In our study, we reconstructed the social network of the Intelligent Communication Laboratory, College of Information Science and Engineering, Ritsumeikan University. For a typical application of the system, we considered situations where a student having troubles with studying particular subjects, such as networking or programming, seeks an advice for her/his study. She/he thus needs to find out who would be the best candidate – a member of the laboratory – to be requested for help.

The modeled community is composed of 43 members. Each member created a profile containing the member’s name, gender, grade, and certain dynamic characteristics, such as personal network data, availability, and trust values calculated via rating objects. Object rating data have been collected from the laboratory members in regard to some 19 categories about classes (e.g. Math), specialized knowledge topics (e.g. Java programming), and the campus life (e.g. Events). The members were asked to select and rate maximum three other members (i.e. the objects), whom they previously requested for help in the given category. By implementing the information-gathering algorithm described in the previous section, the system can then choose the “best” adviser in the category, as it is socially recognized in the community.

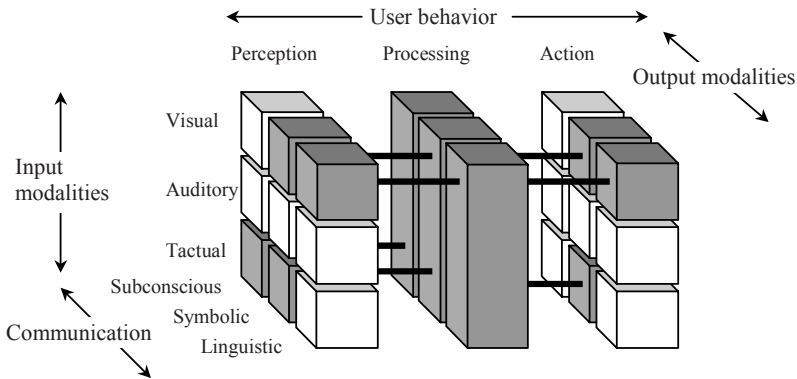


Fig. 2. Exploration of the social space with the multimodal interface

3.1 Multimodal Interface and Interaction

The proposed system has a multimodal interface to facilitate the system-user interaction and increase the efficiency of the navigation process. In the developed prototype, the user receives information about the community not only in a visual form, but also from physically sensing objects and areas in the virtualized social space. For the latter, the user manipulates a haptic device – force display PHANTOM [8]. As the force display reproduces the reaction force, the user receives additional information about the friendliness and socializability “structure” of the community. In addition to representing the community’s physical disposition in the “traditional” 3D virtual reality, the interface then delivers relevant social and personal information via the tactile

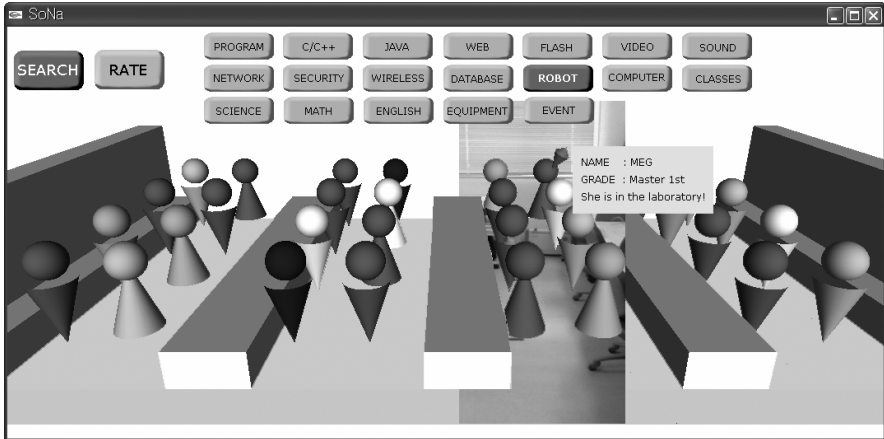


Fig. 3. Screen-shot of the prototype interface

channel by imposing force fields on the 3D graphics space. Fig. 2 illustrates modalities of the user-system interactions for the developed prototype (in the figure, gray blocks indicate active channels).

Fig. 3 is a screen-shot of the prototype’s interface, where the objects – the members of the community – are represented with cone and sphere polygons. The objects are placed in a virtual space partly reconstructing the laboratory settings with relevant photographic images interactively displayed for locations currently under exploration. (In the screen-shot the “active” area around the cursor is the two lines – the 3rd and 4th from the right – of students highlighted with the photographic image window in the background.) The shape of the displayed objects depends on the member’s gender, and the color – on the grade. When the user “touches” an object with the haptic interface pointer, the corresponding user profile appears. If the user touches the “Search” button on the screen, new 19 buttons appear and are used to specify a category of the user’s request. Once a category is selected, the agent of the user attempts to find the best advisers, using the agent network. The best three members are proposed by the recommender system, and the chosen advisers’ (i.e. objects) are indicated with brighter color tones, while scalar force-fields are reproduced with the haptic device to assist the user’s navigation in the social space created by the members present in the displayed physical space.

3.2 Haptic Force Modeling

The described prototype was implemented in C++ programming language with its haptic environment built on top of the SmartCollision Studio™ commercial software package. The latter package provides penetration depth calculation in real-time with realistic friction and elastic force modeling during collisions of the haptic interface pointer (HIP) and (visualized) geometrical objects. The stiffness coefficient was associated with personal “friendliness” of the displayed member, e.g. “rigid” stands for an “unfriendly” subject. After a short training, the participants of our experiments (see Section 4) could easily discriminate “soft”, “average,” and “rigid” subjects for the

stiffness coefficients of 75, 200, and 350 N/m, respectively. The friction coefficient was associated with the member's socializability (higher friction – better socializability – easier to perceive/harder to miss or overlook). Usually, the feedback force is set to zero when there is no collision of the HIP with objects. However, this method does not allow for discriminating between perceptions of various groups of “social objects” and of the “free” space. For the entire social space exploration, we introduced a new approach based on scalar viscosity fields set around “non-friendly” groups of objects, and attracting force fields – around “friendly” groups.

The dynamic model of the method is described as follows. Consider first a point of mass m in a viscosity field λ . Assume that the point is loaded by external “attraction” and driving forces, F_a and $-F_h$, respectively. The point dynamics is then defined with the following equations:

$$m\ddot{r} + \lambda(r - r_1)\dot{r} = F_a(r - r_0) - F_h(t), F_h(t) = k_h\Delta r(t) + b_h\Delta \frac{dr(t)}{dt}, \quad (4)$$

where $r = (x, y, z)^T$ is the point radius-vector. The driving force is opposite to the haptic feedback force F_h , which is calculated in real-time by the standard “spring-damper” model [2,3] using the PHANToM coordinate input. In equations (4), $\Delta r(t)$ is a vector from the current HIP position to the mass point, k_h and b_h are coefficients of the “spring-damper” model. For simplicity, we used only one “attraction” pole at position r_0 , and calculated F_a as the force in the direction to r_0 and proportional to the distance between r and r_0 . Likewise, we selected only one focus of “unfriendliness” at r_1 , and calculated an isotropic scalar viscosity field which depends on the distance to the focus of “unfriendliness” as $\lambda(r-r_1)=a/(1+(r-r_1)^2)$, where a and b are some constants. In the beginning of haptic interaction, it is assumed that the HIP and the mass points coincide and then, the corresponding differential equation (4) is numerically solved by a fourth-order Runge-Kutta method, using the real-time control function $F_h(t)$. In our experiments, parameters a and b were set to provide the viscosity from 1 kg/s to 15 kg/s inside the working PHANToM space.

The physical values of the coefficients (stiffness, friction, and viscosity) for the haptic model were set as linearly proportional to the social characteristics obtained from profiles of the members. The proportionality constants were adjusted to yield smooth haptic feedback, which obviously makes it difficult to move around an “unfriendliness” pole, and which “guides” in the direction to the focus of “attraction”. The personal friendliness is set proportional to the total number of people declaring the member as “friend,” and the socializability – to the number of contacts in the member's personal network. The pole location of the viscosity field is defined by the minimum of the whole group (i.e. summed up) friendliness, while the maximum group friendliness yields the attraction pole location. Model (4) with one attracting pole and a central radial viscosity field provides subjectively a good intuitive reinforcement guidance in the social space. It was found experimentally that haptic guidance becomes ambiguous when the number of poles is more than two. Therefore, only the global minimum and maximum of the group friendliness function were used.

4 Pilot Study

The developed prototype was installed on an ordinary desktop computer (with the haptic device connected) in the Intelligent Communication Laboratory. The laboratory members were asked to review and possibly update their profiles in the system database at least once a week during the spring semester when the pilot study took place. The laboratory is also equipped with a semi-automatic system monitoring the current location of each member – this data was used to automatically update the location dynamic parameter of the agents in the prototype’s agent network. Two groups of 3rd-year students newly assigned to the laboratory (the students are expected to join the laboratory at the beginning of the next academic year), who are generally unfamiliar with (and not yet members of) the community, were formed on a random basis: 7 students (2 females and 5 males) in group I (the control) and 6 students (1 female and 5 mails) in group II. Both groups were asked to file their experiences of seeking advices or help from members of the laboratory for the period of 1 month in the middle of the spring semester. (It is regular practice and is promoted by the teachers at universities in Japan that younger students enquire their older and, assumingly, more experienced and knowledgeable colleagues for help in studying “difficult subjects.”) Group I did not use the prototype, while group II was familiarized with the user interface of the prototype but received no explanations about the specific “meaning” of the reaction force feedback in the haptic environment. The students in both groups were asked to file only the cases when no cross-group communication occurred, and the students of group II were requested to always use the prototype to select advisers. 27 experiences were reported by group I, and 24 – by group II. The reported success rate (as it was subjectively judged by the involved students) stood at 67% in group II and 41% in group I.

5 Concluding Remarks

The experimental results obtained in the pilot study suggest that the proposed system can be a useful tool for assisting human navigation in unknown social environments and for increasing the efficiency of information search in such environments. It is understood, however, that while the developed system is a working prototype that can be used in practice “as is,” at least some of the design solutions implemented in the multimodal interface are rather arbitrary. Larger scale and more elaborated experiments need to be conducted to justify the choice of specific parameters of the social network, which are used to construct the social space in the virtual environments. It was observed by the authors that it is the trust value dynamics that strongly affects the community’s communication patterns and navigation in the social space. It remains, however, unclear if and how the trust value parameter would directly be used to reconstruct the social space.

A deficiency in the authors’ current understanding of the developed system utility is the role of the subconscious tactile interaction channel. As an audio interaction channel would naturally be added to the multimodal interface (that is, in fact, part of the authors’ plans for future work), separate experiments should be conducted to analyze how the interactions in different modalities affect the navigation process.

The main contribution of the presented work, as seen by the authors, is the original concept of the social navigation support service and the design of the information system to realize this service. Specific design details may and will be changed, however. In the next version of the prototype, we plan to significantly increase the size of the agent network and the number of knowledge categories supported. Attempts will also be made to improve the scenes reproduced in the 3D graphics virtual reality along with the haptic image of the social space.

References

1. Derene, G.: How Social Networking Could Kill Web Search as We Know It. *Popular Mechanics* (2008), <http://www.popularmechanics.com/technology/industry/4259135.html>
2. Goncharenko, I., Svinin, M., Kanou, Y., Hosoe, S.: Predictability of Rest-to-Rest Movements in Haptic Environments with 3D Constraints. *Robotics and Mechatronics* 18(4), 458–466 (2006)
3. Goncharenko, I., Svinin, M., Kanou, Y., Hosoe, S.: Skilful Motion Planning with Self- and Reinforcement Learning in Dynamic Virtual Environments. In: Proc. of the 4th INTUITION International Conference, October 4-5, Athens, Greece, pp. 237–238 (2007)
4. Kalman, Y.M., Rafaeli, S.: Modulating synchronicity in computer mediated communication, In: Proc. of the 2007 Conference of International Communication Association, May 24-28, 2007, San-Francisco, USA (2007), <http://www.kalmans.com/synchasynchICAsubmit.pdf>
5. Kryssanov, V.V., Okabe, M., Kakusho, K., Minoh, M.: Communication of Social Agents and the Digital City – A Semiotic Perspective. In: Tanabe, M., van den Besselaar, P., Ishida, T. (eds.) *Digital Cities 2001*. LNCS, vol. 2362, pp. 56–70. Springer, Heidelberg (2002)
6. Lefebvre, H.: *The Production of Space*. Translated by Donald Nicholson-Smith. Blackwell, Oxford (1994) (the original published in 1974)
7. Monge, P., Contractor, N.: *Theories of Communication in Networks*. Oxford University Press, Oxford (2003)
8. SensAble Technologies Inc, <http://www.sensable.com/>
9. Spence, R.: A framework for navigation. *Int. J. of Human-Computer Studies* 51(5), 919–945 (1999)
10. Walter, F.E., Battiston, S., Schweitzer, F.: A Model of a Trust-based Recommender System on a Social Network. *Autonomous Agents and Multi-Agent Systems* 16(1), 57–74 (2008)

Towards Haptic Performance Analysis Using K-Metrics

Richard Hall¹, Hemang Rathod¹, Mauro Maiorca¹, Ioanna Ioannou¹,
Edmund Kazmierczak², Stephen O Leary³, and Peter Harris⁴

¹ Melbourne University Virtual Environment for Simulation

² Department of Computer Science and Software Engineering

³ Department of Otolaryngology

⁴ Biomedical Multimedia Unit, Faculty of Medicine, Dentistry and Health Sciences,
University of Melbourne, Victoria, Australia

Abstract. It is desirable to automatically classify data samples for the assessment of quantitative performance of users of haptic devices as the haptic data volume may be much higher than is feasible to manually annotate. In this paper we compare the use of three k-metrics for automated classification of human motion: cosine, extrinsic curvature and symmetric centroid deviation. Such classification algorithms make predictions about data attributes, whose quality we assess via three mathematical methods of comparison: root mean square deviation, sensitivity error and entropy correlation coefficient. Our assessment suggests that k-cosine might be more promising at analysing haptic motion than our two other metrics.

Keywords: Haptic performance analysis, motion classification.

1 Introduction

The number of devices that interface between user and computer using different sensory modalities are continuously increasing as is their applications, including gaming entertainment, intelligent impairment assistance, and training. Haptics enabled virtual reality simulators are becoming a key adjunct to traditional methods of surgical training, providing trainee surgeons with the *'feel'* of a procedure as well as its visual aspects by combining visual, haptic, and auditory interfaces. Within our domain of interest there exist a number of visuo-haptic simulators for temporal bone surgery [1,2,3,4,5,6,7], which purport to imitate real-world interactions between surgical instruments and temporal bone anatomy whilst providing appropriate visuo-haptic feedback to surgical trainees.

Aside from developing simulations of surgical procedures, researchers have also developed ways to evaluate trainee performance on these simulations [8]. The simulators by Sewell, Morris et.al. [2,9] produce a number of metrics that facilitate evaluation of the trainee's ability to remove the correct amount of bone using the correct drilling tools and drill speed while drilling at a *safe* distance from sensitive anatomical structures. However, we know of no prior work in analysing haptic motion in the way we describe in this paper.

In temporal bone surgery, the technique used to remove bone is to sweep the surgical drill across the bone in a motion that is often referred to as a *stroke*. The choice of stroke is critical to the success of a procedure. For example, an expert will shorten their strokes as they drill in proximity of a sensitive anatomical structure such as the facial nerve, and lengthen their strokes when they are certain that are in an area in which they can work quickly.

This paper evaluates three methods for the automatic annotation of streams of haptic data into strokes. It is organised as follows. In Section 2 we describe the factorial design for our experiment and in Section 3 we discuss the specific metrics that we use to automatically annotate strokes. In Section 4 we discuss several methods to assess the predictive performance of our algorithms. In Section 5 we apply the metrics to ten data, assess the metrics, and present the results of this comparison. Subsequently there is a brief discussion and conclusion.

2 Experimental Design

What exactly is a stroke? Our haptic tool constantly streams out positional information, which is essentially a time series T of three dimensional points P :

$$T = \{P_j = (x_j, y_j, z_j) | j = 1, 2, 3, \dots, m\}; \quad (1)$$

where P_j denotes a *temporal* neighbour of P_{j-1} and P_{j+1} , and (x_j, y_j, z_j) is the Cartesian coordinate of P_j . We would consider the vector from P_j to P_{j+1} to be a micro-stroke. On the other hand, it is also possible to describe T as a set of subsets:

$$T = \{S_i = \{(x_{ij}, y_{ij}, z_{ij}) | j = 1, 2, 3, \dots, m\} | i = 1, 2, 3, \dots, n\} \quad (2)$$

where S_i is a stroke. As a sub-sequence of T it also has temporal relations to other strokes; it follows stroke S_{i-1} and precedes stroke S_{i+1} . To identify strokes, we require a classification function F_c (see Equation 3) for points.

$$F_c(X_{P_b}) = \begin{cases} 0 : & P_j \in S_i, & P_{j+1} \in S_i \\ 1 : & P_j \in S_i, & P_{j+1} \in S_{i+1} \end{cases} \quad (3)$$

Given the unpredictability of the deterministic relationship between strokes, we used two statistical thresholds for F_c , $ST_1 = \mu + \sigma$ (weak filter) and $ST_2 = \mu + 2\sigma$ (strong filter), assuming a normal distribution [10]. The reason that the parameter for Equation 3 is not simply P_j is because an isolated point in 3D space doesn't contain enough information to determine to which stroke it belongs. So we need an information-rich attribute X_{P_j} for each P_j ; one which has been heavily used in pattern recognition and artificial vision is high curvature [11, 12]. We chose to use a mature approach for measuring curvature known as k-cosine [13] (still popular due to its simplicity [14]) and two other methods that are similar in conception. The basic idea of this method is to take any point P_j then pick only two points, P_{j-k} and P_{j+k} (k points away from P_j). These three points are

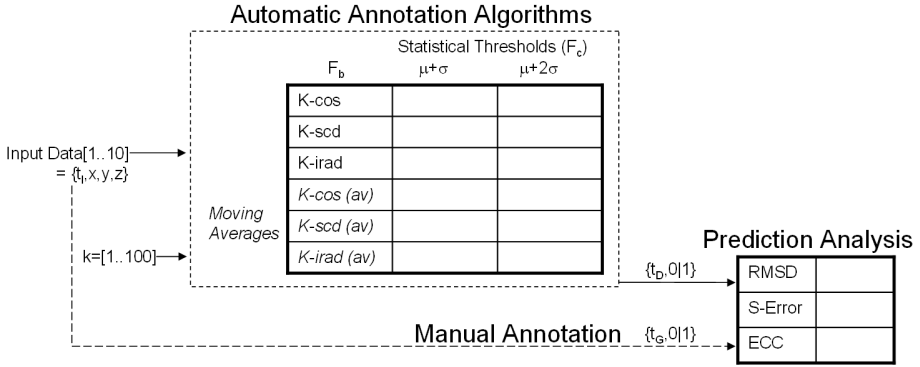


Fig. 1. Experimental Design

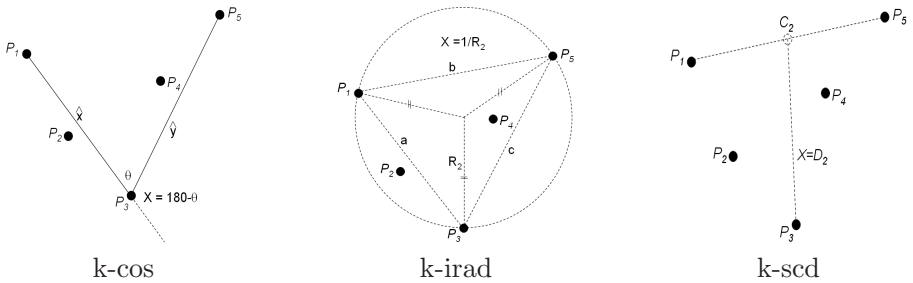
input to a boundary characterisation function F_b which produces the required information-rich attribute X_{P_j} :

$$X_{P_j} = F_b(P_{j-k}, P_j, P_{j+k}) \tag{4}$$

We use a factorial design (see Figure III) for our experiment. We input 10 data sets (seven benchmark images and three realistic haptic data streams) which are then processed by F_b (our three k-metrics, with and without moving average [-k..k] ala [15]), then filtered using F_c (statistical thresholding). The predictive power of these automatic annotation algorithms is then analysed by several methods of mathematical comparison with a manually constructed ground truth, which we explore later in this paper.

3 K-Metrics

In this section we describe the three k-metrics for F_b we used in our investigation: cosine (k-cos), extrinsic curvature (k-irad), and symmetric centroid deviation (k-scd) (shown below for k=2)



Since all of our methods were based on the general idea of k-cosine (k-cos), we describe it first. In this method, the boundary characterisation function F_b

is related to the cosine of the angle between the vectors (see Equation [5](#) [16](#)). The metric $X_{P_b} = 180 - \theta$. It accounts for point co-linearity: if $P_a = P_c$ then $X_{P_b} = 180$, otherwise if P_b lies directly between P_a and P_c then $X_{P_b} = 0$.

$$\cos\theta = \frac{\hat{x} \cdot \hat{y}}{|x||y|} \quad (5)$$

Our second method, k-extrinsic curvature (k-irad), does more than consider the angle between three points - it also factors in the distance between the three points. The so-called *circumradius* (see Equation [6](#) [17](#)) can be trivially calculated between any three points that are non-colinear as shown below (also for $k=2$). The boundary characterisation function F_b is the extrinsic curvature (the inverse of the radius). This method accounts for point co-linearity in a slightly different way than k-cosine. If $P_a = P_c$ then the radius is half the distance from P_a to P_b , which is unbounded (as opposed to k-cosine). On the other hand, if P_b lies directly between P_a and P_c then the denominator of Equation [6](#) is zero so our algorithm uses an early check to set $X_{P_b} = 0$.

$$R = \frac{abc}{\sqrt{(a+b+c)(b+c-a)(c+a-b)(a+b-c)}} \quad (6)$$

Our third metric, k-symmetric centroid deviation (k-scd), simply calculates the centroid C_k between P_a and P_c then X_{P_b} is simply the distance between the centroid and P_b (see Equation [7](#)). This method accounts for co-linearity similarly to k-extrinsic curvature. If $P_a = P_c$ then X_{P_b} is simply the distance from P_a to P_b . Else if P_b lies directly between P_a and P_c then $C_k = P_b$ so $X_{P_b} = 0$.

$$X = \sqrt{(c_k - p_i)^2} \quad (7)$$

In this section we described the three basic boundary characterisation functions F_b that will be investigated both with and without a moving average.

4 Assessing Predictive Performance

Which of our six methods makes better predictions than another? There are three popular mathematical methods for comparing predicted values with observed values: root mean squared deviation (RMSD); true-positive ratios; and mutual information measures. RMSD (see Equation [8](#)) is a widely-used statistical measure of the difference between values predicted by a model D and the nearest values actually observed from the thing being modeled or estimated G [18](#).

$$RMSD(D, G) = \frac{\sqrt{\sum_{i=1}^n (x_{Di} - x_{Gi})^2}}{n} \quad (8)$$

However, perspective differences exist, depending on the parameters of this equation. For example, if the algorithm predicted every point as a stroke boundary, the RMSD from the perspective of the ground truth would be zero, but from the perspective of the prediction it would be very large. Thus we use:

$$\varepsilon = \max \{RMSD(D, G), RMSD(G, D)\} \quad (9)$$

RMSD is certainly a useful measure, a low RMSD means that predictions are accurate. However a high RMSD is ambiguous; it can mean that predictions generally are bad, or that predictions generally are reasonable but can be skewed by outliers. So, rather than just looking at averages, we considered other statistical measures relating accurate to inaccurate predictions (e.g. Equation 10 [19]).

$$\text{Sensitivity} = \frac{\text{No. of true positives}}{\text{No. of true positives} + \text{No. of false negatives}} \quad (10)$$

However we avoid such measures directly because they rely on choosing on a distance ϵ from the ground truth within which a prediction is considered true. Instead, with Equation 11 we calculate the maximum (ϵ) between all ground truth points and the corresponding closest data point which would give sensitivity = 1. Calculating the sensitivity error is the same as calculating the furthest outlier which would impact the RMSD, so a low sensitivity error means greater prediction stability.

$$\text{Sensitivity Error} = \max\{\forall G \forall D \min\{P_G(t) - P_D(t)\}\} \quad (11)$$

Finally, there are several information-theoretic methods able to compare how dependent G is on D considered as distributions. In the medical imaging literature a very popular method for comparing dependence is mutual information [20], we prefer a related measure called entropy correlation coefficient (ECC) [21] because ECC=0 means no dependence and ECC=1 means full dependence. In order to construct ECC we first find the entropy of D and G singly (see Equation 12), then find the joint entropy (see Equation 13).

$$H(D) = - \sum_D p_D \log_2 p_D \quad (12)$$

$$H(D, G) = - \sum_{D, G} p_{D, G} \log_2 p_{D, G} \quad (13)$$

Subsequently, we create a ratio measure relating the single and joint entropies called normalised mutual information (see Equation 14 [22]) then scale NMI into the range 0 to 1.

$$NMI(D, G) = \frac{H(D) + H(G)}{H(D, G)} \quad (14)$$

$$ECC(D, G) = 2 - \frac{2}{NMI(D, G)} \quad (15)$$

Thus we have three methods for comparing algorithms, RMSD, sensitivity error and entropy correlation co-efficient.

5 Results

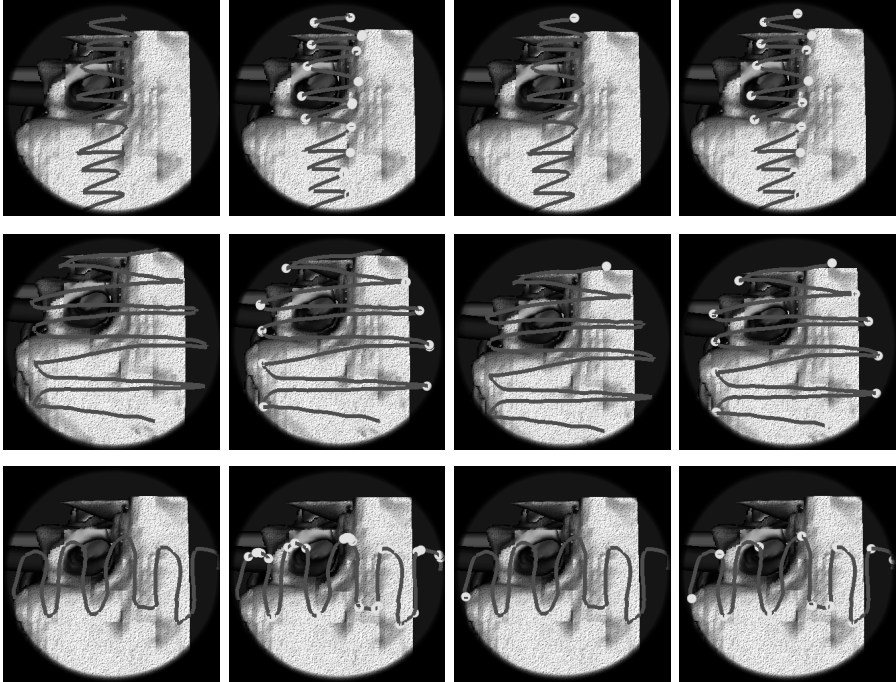
We test our algorithms on two data sets (3D haptic and 2D benchmark) with our six k-metrics: K is distance from point of interest and ST is filter strength.

The *haptic* data set consists of three stroke types: fast (magnified), slow and wavy (non-overlapping thus amenable to screen-grab).

Haptic	K	ST	RMSD		S-Error		ECC	
			Mean	StD	Mean	StD	Mean	StD
k-cos	6	1	1.378	0.656	5.333	2.517	0.259	0.255
k-cos (av)	8	2	1.403	0.6	7	3.606	0.293	0.285
k-scd	4	2	6.367	0.777	85	19.16	0.01	0.004
k-scd (av)	5	2	6.817	1.268	94	26.85	0.009	0.005
k-irad	3	2	1.707	1.005	12.67	15.14	0.098	0.086
k-irad(av)	8	2	1.706	0.455	9.667	9.292	0.249	0.171

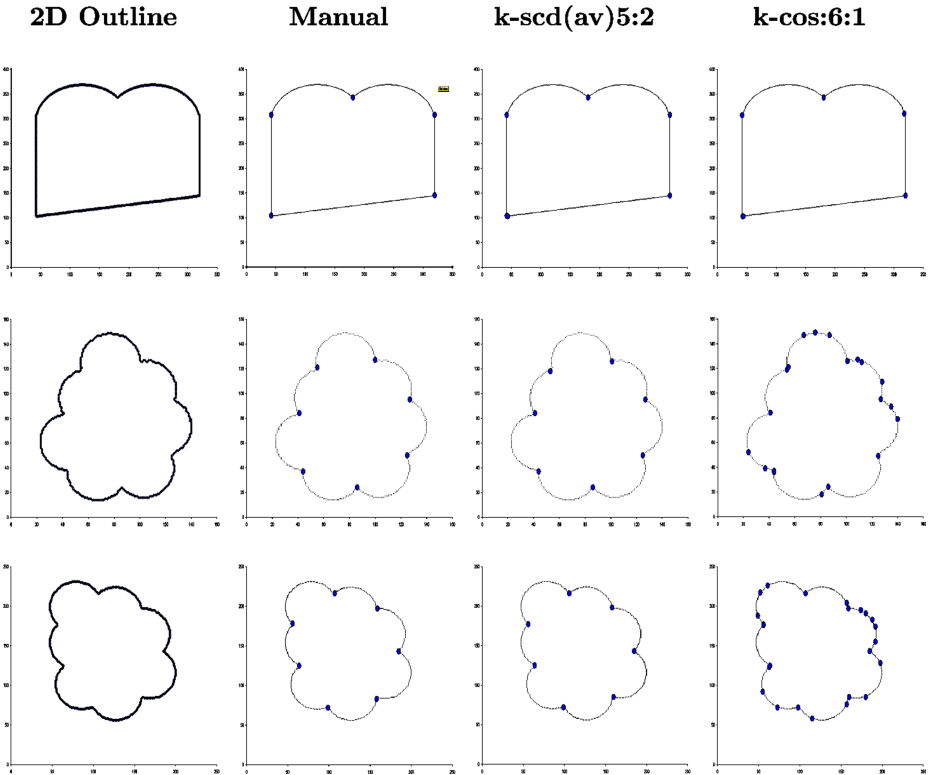
The four columns below show: hand motion, ground truth, then two example predictions made by the best performing methods for both data.

Haptic Path **Manual** **k-scd(av)5,2** **k-cos:6:1**

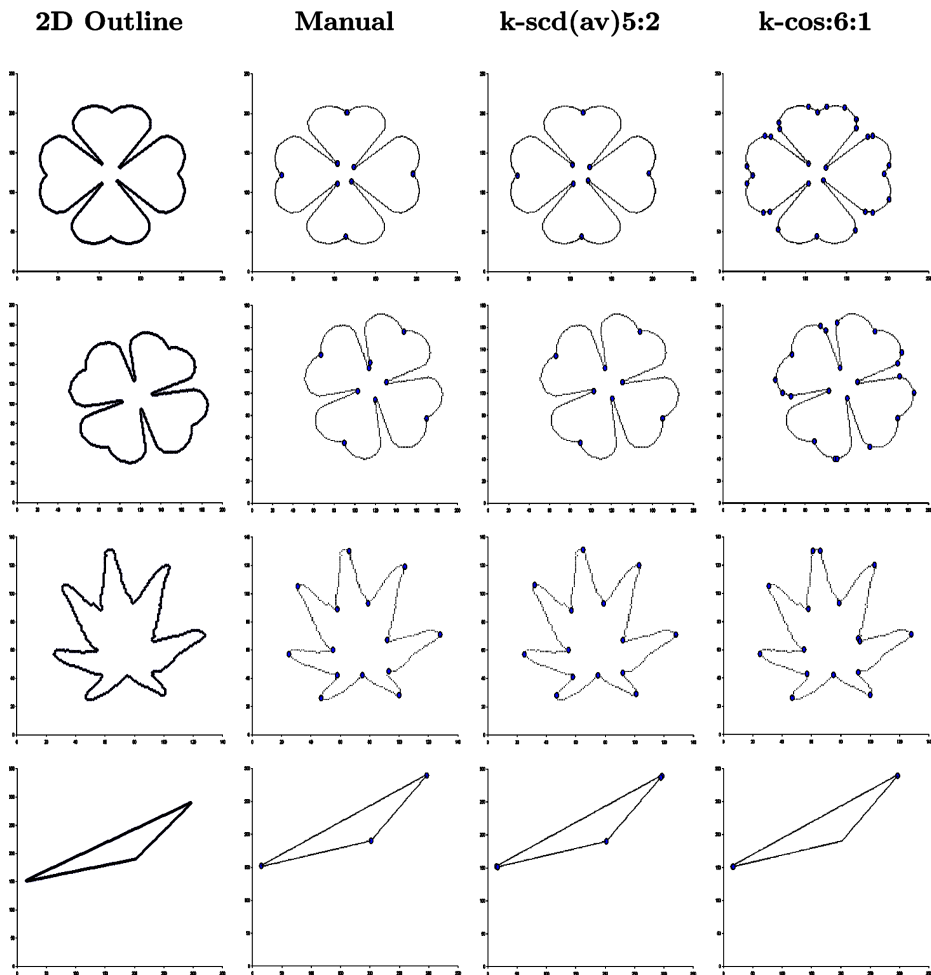


The *2D benchmark* data set consists of seven images used for testing corner detection methods [14]. These images were adapted to our temporal domain in the following way. First, the images were morphologically eroded by one pixel to shrink the original image I_1 by one pixel to make a new image I_2 . Then I_2 was subtracted from I_1 to produce the image outline I_3 . We then developed a simple edge crawling algorithm to convert I_3 into a time series of 2D points (which is exactly the same as T from Section 4 but all $Z=0$).

	K	ST	RMSD		S-Error		ECC	
			Mean	StD	Mean	StD	Mean	StD
2D								
k-cos	6	1	2.535	2.631	1.571	1.272	0.361	0.167
k-cos (av)	8	2	2.751	2.784	27.71	43.39	0.478	0.126
k-scd	4	2	2.448	2.757	2.571	1.902	0.408	0.217
k-scd (av)	5	2	2.454	2.737	2.429	2.07	0.475	0.232
k-irad	3	2	2.493	1.488	21.14	35.26	0.025	0.059
k-irad(av)	8	2	2.784	1.456	23.43	35.19	0.002	0.001



The algorithm that made the best predictions of the haptic data on average was k-cos, with a k value of 6 and a statistical threshold of $\mu + \sigma$, which produced a good average RMSD of 1.378 and a competitive ECC of 0.255. Visual comparison of the manually constructed haptic ground truth with k-cos:6:1 shows good face validity. With respect to the 2D benchmark images, k-cos:6:1 had comparatively poor face validity, often generating high numbers of false positives. K-scd:4:2 and k-scd(av):5:2 both had low RMSD on the 2D benchmark images (2.448 and 2.454 respectively), which we considered a tie. However k-scd(av):5:2 had a higher entropy correlation coefficient (0.475 as opposed to 0.408), so we considered it to be superior in the 2D case. Visual comparison of the manually



constructed ground truth for the 2D benchmark images shows good face validity for $k\text{-scd}(\text{av}):5:2$. Note that in this analysis we ranked RMSD above sensitivity error and ECC because of the greater meaningfulness of the magnitude of this single measure.

6 Conclusion

This paper presented and compared six k -metrics for classifying haptic data, in our case, motion recorded from a user of a virtual drill in a haptic-enabled temporal bone drilling simulation. Our preliminary results suggested that $k\text{-cos}$ might be a competitive method for the automatic annotation of haptic motion in terms of strokes. Thus, this metric seems well suited for implementation in haptic simulations which require quantification of the relation between hand motion and performance. However, other experimental results showed that $k\text{-symmetric}$

centroid deviation was superior for the 2D benchmark images, thus perhaps for other two dimensional and relatively smooth data like sound waveform sampling and classification it would be preferable to use this method.

The ability to correctly identify strokes is beneficial for several reasons. Firstly, it allows hypotheses about differences between experts and novices with respect to stroke differences to be investigated within thousands of lines of haptic tool tracking data which would be infeasible to annotate manually. If such differences can be modelled, it should be possible to automatically assess surgical trainee performance, both for giving trainees feedback (realtime and otherwise) and also as part of a suite of trainee assessment tools.

Acknowledgments

We thank Matthew Hutchins and Chris Gunn from the CSIRO ICT Centre who initially worked on the temporal bone simulator with which we are continuing. We also thank Les Kitchen from Melbourne University whose comments and suggestions on a draft of this paper were invaluable.

References

1. Hutchins, M.A., O'Leary, S., Stevenson, D., Gunn, C., Krumpolz, A.: A networked haptic virtual environment for teaching temporal bone surgery. In: *MMVR XIII*, pp. 204–207. IOS Press, Amsterdam (2005)
2. Morris, D., Sewell, C., Barbagli, F., Salisbury, K., Blevins, N.H., Girod, S.: Visuo-haptic simulation of bone surgery for training and evaluation. *IEEE Comput. Graph. Appl.* 26(6), 48–57 (2006)
3. Rasmussen, M., Mason, T.P., Millman, A., Evenhouse, R., Sandin, D.J.: The virtual temporal bone, a tele-immersive educational environment. *Future Gen. Comp. Sys.* 14(1-2), 125–130 (1998)
4. Wiet, G.J., Bryan, J., Dodson, E., Sessanna, D., Streadney, D., Schmalbrock, P., Welling, B.: Virtual temporal bone dissection simulation. In: *MMVR 2000*. IOS Press, Amsterdam (2000)
5. John, N.W., Thacker, N., Pokric, M., Jackson, A., Zanetti, G., Gobbetti, E., Giachetti, A., Stone, R., Campos, J., Emmen, A., Schwerdtner, A., Neri, E., Franceschini, S., Rubio, F.: An integrated simulator for surgery of the petrous bone. In: *MMVR 2001*. IOS Press, Amsterdam (2001)
6. Agus, M., Giachetti, A., Gobbetti, E., Zanetti, G., Zorcolo, A., John, N.W., Stone, R.J.: Mastoidectomy simulation with combined visual and haptic feedback. In: *MMVR 2002*. IOS Press, Amsterdam (2002)
7. Pflessner, B., Petersik, A., Tiede, U., Hohne, K., Leuwer, R.: Volume cutting for virtual petrous bone surgery. *Computer Aided Surgery* 7(2), 74–83 (2002)
8. Agus, M., Giachetti, A., Gobbetti, E., Zanetti, G., Zorcolo, A.: Tracking the movement of surgical tools in a virtual temporal bone dissection simulator. In: Ayache, N., Delingette, H. (eds.) *IS4TM 2003*. LNCS, vol. 2673. pp. 1004–1011. Springer, Heidelberg (2003)
9. Sewell, C., Morris, D., Blevins, N.H., Agrawal, S., Dutta, S., Barbagli, F., Salisbury, K.: Validating metrics for a mastoidectomy simulator. In: *MMVR XV*. IOS Press, Amsterdam (2007)

10. Kenney, J.F., Keeping, E.S.: Mathematics of Statistics Pt - 1, 3rd edn. Van Nostrand (1962)
11. Inesta, J.M., Buendia, M., Sarti, M.A.: Reliable polygonal approximations of imaged real objects through dominant point detection. *Pattern Recogn. Lett.* 31, 685–697 (1998)
12. Attneave, F.: Informational aspects of visual perception. *Psychological Review* 61, 183–193 (1954)
13. Rosenfeld, A., Johnson, E.: Angle detection on digital curves. *IEEE Trans. Comput.* 22(7), 875–878 (1973)
14. Sun, T.H., Lo, C.C., Yu, P.S., Tien, F.C.: Boundary-based corner detection using k-cosine. *Systems, Man and Cybernetics*. In: ISIC. IEEE International Conference on (October 7-10), pp. 1106–1111 (2007)
15. Rosenfeld, A., Weszka, J.S.: An improved method of angle detection on digital curves. *IEEE Trans. Comput.* 24(9), 940–941 (1975)
16. Arfken, G.: *Scalar or Dot Product*, 3rd edn. Academic Press, London (1985)
17. Johnson, R.A.: *Modern Geometry: An Elementary Treatise on the Geometry of the Triangle and the Circle*. Houghton Mifflin (1929)
18. Maiorov, V.N., Crippen, G.M.: Significance of root-mean-square deviation in comparing three-dimensional structures of globular proteins. *J Mol Biol* 235(2), 625–634 (1994)
19. Kanji, G.K.: *100 Statistical Tests*. SAGE Publications, Thousand Oaks (1999)
20. Pluim, J.P., Maintz, J.B., Viergever, M.A.: Mutual-information-based registration of medical images: a survey. *IEEE Transactions on Medical Imaging* 22, 986–1004 (2003)
21. Collignon, A., Maes, F., Vandermeulen, D., Marchal, G., Suetens, P.: Multimodality image registration by maximization of mutual information. *Medical Imaging, IEEE Transactions on* 16(2), 187–198 (1997)
22. Studholme, C., Hill, D., Hawkes, D.: An overlap invariant entropy measure of 3d medical image alignment. *Pattern Recognition* 1(32), 71–86 (1999)

Multimodal Interaction: Real Context Studies on Mobile Digital Artefacts

Tiago Reis, Marco de Sá, and Luís Carriço

LaSIGE, Faculdade de Ciências, Universidade de Lisboa
treis@lasige.di.fc.ul.pt, {marcodesa, lmc}@di.fc.ul.pt

Abstract. The way users interact with mobile applications varies according to the context where they are. We conducted a study where users had to manipulate a multimodal questionnaire in 4 different contexts (home, park, subway and driving), considering different variables (lighting, noise, position, movement, type of content, number of people surrounding the user and time constraints) that affect interaction. This study aimed at understanding the effect of the context variables in users' choices regarding the interaction modalities available (voice, gestures, etc). We describe the results of our study, eliciting situations where users adopted specific modalities and the reasons for that. Accordingly, we draw conclusions on users' preferences regarding interaction modalities on real life contexts.

Keywords: Multimodal Interaction, Mobile Devices, Studies in Real Contexts.

1 Introduction and Background

Multimodal interaction is a characteristic of everyday human activities and communications, in which we speak, listen, look, make gestures, write, draw, touch and point, alternatively or at the same time in order to achieve an objective. Considering the human perceptual channels [4] through the inclusion of elements of natural human behavior and communication on human-computer interfaces is the main goal of multimodal interaction. Multimodal interfaces can improve accessibility for different users and usage contexts, advance performance stability, robustness, expressive power, and efficiency of mobile activities [5, 6]. Recently, concerning the special needs of several groups of users, many researchers have focused on the design and development of universally accessible systems. These consider: impairments [1] and various usage context variables [2, 3]. Design approaches such as “Inclusive Design” or “Design for all” enhance the usability of the applications and consider the existence of multiple interaction modalities, providing accessibility and support to impaired users and enabling non-impaired users to interact with applications in suboptimal conditions [1]. This kind of interaction has been explored from different points of view [1, 7, 8, 9]. (e.g. support to impaired users, support to non-impaired users in suboptimal situations, augmentation of unimodal activities, games, multimedia applications, etc.). The interaction modalities included on a multimodal system can be used either in a complementary way (to supplement the other modalities), in a redundant manner (to provide the same information through more than one modality), or as an alternative to the

other modalities (to provide the same information through a different modality) [10]. The use of non-conventional interaction modalities becomes crucial when concerning human-computer interaction for users with special needs (e.g. impaired users). In these cases the objective is not to complement the existing modalities of a system with new ones but to fully replace them with adequate ones [11, 12].

On another strand, researchers have focused their work on understanding how to overcome the limitations introduced by the tiny screens available on mobile devices. Sound and gestures have been used to augment or replace the conventional device interfaces [14, 15].

Two approaches are considered regarding the evaluation of multimodal mobile applications: laboratory and field evaluation. Both can consider logging of users' activities, filming and questionnaires. Some of the studies reported on the available bibliography addressed the gathering of knowledge about an already deployed device or application by logging users and interviewing them about their use [16, 17]. Other studies consider the evaluation stage [14, 18, 19, 20], and two specific studies compare field and laboratory evaluation [6, 21], suggesting that multimodal mobile applications should be evaluated and studied on the field, in real contexts, under real constraints.

We conducted a study where users had to manipulate a mobile multimodal artefact (in this case: a questionnaire) in 4 different contexts (home, park, subway, driving) considering different context variables (lighting, noise, position, movement, type of content, number of persons surrounding the user and time constraints). The study aimed at understanding the effect of the mentioned context variables in users' choices regarding the interaction modalities available on our tool (e.g. voice, gestures, touch screen, keypad). In this paper, we present an overview on an artefact creation tool and artefact analysis tool. We fully describe a multimodal mobile artefact manipulation tool, which utilization was studied in different contexts; we present the case study and discuss the lessons learned.

2 Mobile Multimodal Artefact Framework

This section provides an overview on a mobile multimodal framework that enables artefact creation, manipulation and analysis [13]. The creation and analysis tools are described below, in order for the reader to understand how the framework is articulated. The manipulation tool, focus of this study, is described in full detail on the next section.

All the tools and corresponding libraries were developed in C# for Microsoft Windows Platforms. Desktop/laptop/tablet and hand-held versions are available. The latter, particularly in the creation tool's case, are simplified versions of the former.

- **Proactive Artefacts:** Artefacts are composed by pages and rules. Pages nest elements of different types (e.g. text/audio/video labels, text/audio multiple choice objects, text/audio/video answer/recording elements). Each element combines visual and audible presentations. In both modes, they include a type-intrinsic counterpart (e.g. drop down menu for the visual form and an audio description on how to interact with it) and an element specific part: corresponding to its content (e.g. the

textual items of a multiple choice element and their matching audio streams). The content can be static and defined when the element is created (e.g. the textual and audible contents of a label), or dynamic and acquired during interaction (e.g. a free text answer or an audio recording element). When only one mode is available (e.g. textual or audible contents), the other is generated whenever it is possible, according to the device's characteristics (through a text to speech or a speech to text engines).

Rules enable the modification of pages, elements and, most notably, navigation (e.g. if answer is YES goto page Y), based on triggers (e.g. X minutes have passed, users requested next page) and conditions (e.g. users answered YES or BLUE) that activate behaviours (e.g. hide element W on page Z, goto page Y).

- **Artefact Creation:** A wizard based creation tool allows users with no programming experience to create fairly elaborated proactive artefacts in three simple steps: 1) creation of the elements (definition of their content and location); 2) definition of the natural page sequence; and, 3) description of rules (triggers, conditions and behaviours).
- **Artefact Analysis:** The analysis tool enables the evaluation and annotation of artefacts and their utilization. It can work as a result viewer, enabling the analyst to evaluate the final results of an artefact, or as a log player, providing the analysis of result evolution and, most notably, of all user interaction. Moreover, this analysis can be accelerated or delayed, according to the analyst's preferences. This tool played a fundamental role on the studies presented on section 4.

3 Artefact Manipulation

A manipulation tool (Fig. 1), allows artefact emulation and the logging of users' activities throughout artefact utilization.

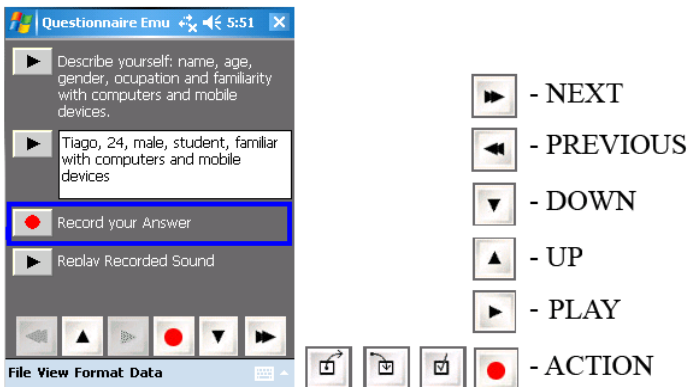


Fig. 1. Artefact Manipulation Tool

Considering devices with touch screen or mouse, data input (e.g. typing text, recording answer, selecting a choice on a list or a value on a track bar) and access to the audible content of the elements can be done directly through the buttons, text boxes, track bars and lists that compose the elements presented on the screen. This is an incomplete interaction modality, since page navigation has to be achieved through any of the other available modalities. These are indirect interaction modalities, in a way that: users must navigate through elements and can only interact with the selected element. These modalities were introduced considering screens without touch technology, movement situations, one hand and no hands interaction [13]. Performance decreases when using indirect interaction modalities but there are no substantial performance differences between the indirect interaction modalities available:

- **Graphical Interaction Bar:** This bar is composed by 6 buttons and it is located on the bottom of the artefact (on Fig. 1). The Next and Previous buttons allow navigation between pages. Navigation between elements of the same page is available through the Up and Down buttons. The Play button plays the audible content of the selected element (shown inside a blue box). The Action button changes according to the selected element, and allows indirect interaction with it: record/stop recording; select a choice; start nested navigation/stop nested navigation. After starting nested navigation the Next and Previous buttons navigate horizontally inside the element (e.g. on a track bar) and the Up and Down buttons navigate vertically inside the element (e.g. on a combo box).

All together, these functionalities support the interaction with the artefacts (except for textual input) and are also accessible through: the device's keypad; gesture recognition; voice recognition; or any combination of the previous.

- **Device Keypad:** This modality maps the over mentioned functionalities to the device's keypad. According to the number of keys available on the device, it can be a complete or incomplete modality. Nevertheless, for keypads with 4 or more keys this is complete interaction modality.
- **Gesture Recognition:** This modality is also built on the mentioned functionalities, mapping them to gestures that are recognized on the device's touch screen. We developed a simple gesture recognition algorithm that allows the recognition of the six different gestures presented bellow (Fig. 2). This algorithm was validated in a laboratory by 10 users, considering different types of interaction (stylus, one hand, two hands) and presenting accuracy rates close to 100%.

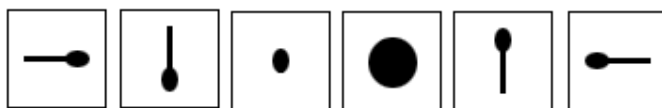


Fig. 2. Gesture interaction – mapping interaction functionalities to gestures. From the left to the right: *Previous*, *Up*, *Play*, *Action*, *Down*, *Next* (see also Fig 1). The small dot represents a tap on the device's screen; the line represents a continuous gesture after the tap; the big dot represents a tap and hold lasting more than 1 and a half second.

- **Voice Recognition:** This modality maps the referred functionalities to voice commands that are recognized by the application: *Previous, Up, Play, Action, Down, Next* (see Fig 1). The set of commands elected for this interaction modality was also validated (in a laboratory, under optimal noise constraints) by 10 users, presenting accuracy rates close to 100%.

Any combination of these interaction modalities can be used, and the users have the possibility of disabling and enabling them according to their preferences. Moreover, this tool can run in different modes: visual, eyes free, hands free, or any combination of the previous:

The visual mode ignores the elements' audible counterpart throughout artefact manipulation and provides interaction through any of the available modalities.

The eyes free mode is directed for situations when users cannot pay much (e.g. walking or running) or any visual attention (e.g. driving) to the device. This mode relies mostly on the indirect interaction modalities (voice recognition, gesture recognition or keypad). When navigating throughout pages, users are audibly informed about the current page. Whilst navigating between elements, users are audibly informed about the type of the selected element and how to interact with it; this information can be skipped using any of the functionalities that support interaction. Direct interaction with elements and the visual counterpart of the artefact are also available, addressing situations when users are in a context that enables them to spare exoradic visual attention to the device.

The hands free mode considers only voice recognition for input and any of the available modalities, or combinations of them, for output.

4 Case Study

The main goal of this study (Fig 3) was to understand which interaction modalities are preferred by users in different contexts and considering different types of information.

Completing a questionnaire is a task that can be done on the move between different contexts. Questionnaires can address different types of information (e.g. private, not private) and contexts may have different characteristics (e.g. noise, lighting, movement status, position, number of persons in the same context, time constraints, etc). Due to the mobile nature of these artefacts and their purpose, we decided to study them on the field, on a set of scenarios selected by the users and our team: home, park, subway and car (driving). This study was conducted in Lisbon (Portugal) and 15 non-impaired users were involved: 8 male, 7 female, ages comprehended between 20 and 35, familiar with computers and mobile phones.



Fig. 3. Study in real contexts – from the left to the right: home, park, subway, driving (2 pictures)

Before conducting the study, the users were taught on how to interact with the artefact manipulation tool through the 5 different interaction modalities (direct interaction with elements, indirect interaction through: keypad, graphical interaction bar, gestures or voice). For each context a group of users was given a questionnaire they had to fill using the artefact manipulation tool. This questionnaire: considered private and non-private information, gathered users' data for statistics, ratings on the application's usability and registered users' opinions about their experience. The questionnaire was composed of 5 open ended questions (that could be answered through textual input or audio recording) and 1 close ended question (question 6, where users had to select 1 out of 6 options). The 6 questions that compose the questionnaire are presented below:

- 1) Describe yourself mentioning: name, age, gender, occupation, familiarity with computers and mobile devices.
- 2) What is your phone number?
- 3) What is your address?
- 4) How have you used this application (including if you used voice, gestures, pen, keypad, virtual keyboard), why have you used it this way and in which situations?
- 5) What were your difficulties during this task?
- 6) How hard was it to perform this task? (choices: impossible, very hard, hard, normal, easy, very easy)

All the tasks performed by the users were logged and filmed. The cross analysis between the utilization logs, movies and questionnaires enabled us to gather the information presented below:

- **Home:** The study conducted in this scenario involved 4 users that filled the same questionnaire twice. The users were sitting down using a TabletPc, on silent and well illuminated environment. The first task had no limitations regarding time. The 4 users chose to fully interact with the questionnaire (navigation between pages, navigation between elements, input on open ended answers and input on closed answers) through voice modalities (recognition and recording). They placed the device on the table and tried to interact with it without using their hands. Three of these users were successfully interacting through voice modalities. The other user had difficulties pronouncing the English words that supported navigation, after realizing he could not use the voice recognition properly, he started: navigating through the graphical interaction bar; interacting directly with the elements on the touch screen; and, performing data input on open ended questions through audio recording; Despite the personal content of some questions, none of the users decided to perform textual input on any of the open ended questions.

Users reported that they tried to explore the modality they found more interesting. We believe this decision was due to the optimal characteristics of this scenario and to the absence of time constraints on the first task. Accordingly, we decided to conduct a second task including a time factor in order for users to try to make an optimal use of the application. When performing this task, users chose direct interaction to perform data input: using the stylus to select multiple choices and to start and stop recording open ended answers. None of the users decided to perform textual input on any of the open ended answers. Navigation between pages was done through voice

recognition and navigation between elements was not used. The mix of modalities appeared only once, in navigation between pages, when the user with pronunciation problems tried to use voice recognition again. The recognition failed two times in a row and the user started navigating through the graphical interaction bar.

For both tasks: 3 of the users considered them easy and 1 very easy.

- **Park:** The study conducted in this scenario involved 5 persons using the Questionnaire Manipulation Tool on PDAs. Users filled the same questionnaire twice: first sitting and then walking, always considering time constraints. The study was conducted during the day, on a park close to a road with constant traffic: the environment was noisy and over illuminated (the reflection of the sun on the screen reduced the visibility of the artefact significantly).

On the first task, the modality adopted to navigate between pages varied: 3 used voice commands, 1 gestures and 1 used the device's keypad. The 3 users navigating through voice commands and the one navigating through gestures justified their choices based on their personal opinion about which was the most interesting interaction modality. Nevertheless, in occasions where the voice recognition failed 2 or 3 times (due to the noise), users tried gesture recognition justifying it with the interest factor mentioned before. The user navigating through the keypad justified his choice by saying: "it's the most practical from my point of view". None of the users considered navigation between elements. The input on open ended answers also varied: 4 users recorded their answers to questions 1 and 3 and typed their answers to questions 2 and 3 on the virtual keyboard. These users justified their choice with the fact that didn't feel comfortable about saying their phone number and address out loud in a place where there were people they didn't knew close by; 1 user, the one using the keypad to perform navigation, embarrassed about speaking out loud to a device on a place where there was more people answered all opened questions through the virtual keyboard. The 5 users chose to perform input on closed answers directly on the elements using the stylus. This task was rated: easy by 4 users and normal by 1.

On the second task, the modalities chosen to navigate between pages and the justifications for those choices were the same as for the first task. Nevertheless, in this task all users considered navigation between elements. The input on closed answers was done, indirectly by the 5 users, through the same modality they had chosen to perform navigation between pages on the first task. The elected modality for input on open ended answers was the audio input, users recorded their answers to all open ended questions, but they all reported that at least the personal questions they wouldn't answer in such situation, because they didn't feel comfortable about saying their phone number and address out loud in a place where there were more people close by and because it was "very difficult and boring" to write on a virtual keyboard while walking. This task was rated: normal by the 5 users.

- **Subway:** This study involved 4 users equipped with PDAs. These, performed the same task twice, first sitting and then standing. The environment was very noisy, well illuminated and presented a large set of movement patterns (e.g. accelerations, breakages, people entering and leaving, etc.).

On the first task the modality adopted to navigate between pages varied: 1 user chose keypad and the other 3 gestures. Users justified their choices based on their

personal opinions about which were the most interesting navigation modalities. They justified the exclusion voice recognition based on their embarrassment of speaking to a device in the subway surrounded by so many people. For the same reason, the input on open ended answers was always done through the virtual keyboard. None of the users considered navigation between elements and all of them chose to perform input on closed answers directly on the elements, using the stylus. This task was rated very easy by the 4 users.

On the second task all the users were holding and interacting with the PDA using only one hand, while they used the other to hold themselves safely in the subway. The modalities adopted to navigate between pages were the same as in the first task and they were justified with the same reason, which also justifies the textual input on answers to open ended questions. Users reported difficulties using only one hand to hold and interact with the device, especially when attempting to perform finger interaction with a virtual keyboard available on touch screens. All users considered navigation between elements and indirect interaction with them, through the same modality they chose to perform navigation between pages. This task was rated hard by the 4 users.

- **Driving:** The study conducted on this scenario involved 2 users that, due to safety concerns of our team, filled a questionnaire while driving in a street with little traffic, without considering time constraints. The environment was noisy and well illuminated. The users were sitting down performing hands and eyes free interaction with the application running on a TabletPc that was placed on the sit next to them. This was the only context in which users chose to use the audible counterpart of the questionnaires. On this task users used voice recognition to navigate between pages, elements, and to indirectly answer closed questions. The input on open ended questions was performed through audio input. These choices were justified based on the fact that users were driving using both hands and that they could not spare almost any visual attention to the device due to its location. Both users rated this task normal, considering its difficulty.
- **Discussion and lessons learned:** Regardless of the inclusion of time constraints on the tasks, users tended to interact with our application through the modalities that interest them the most from a technological and innovational point of view. When time constraints were introduced users still manifested this tendency but only for activities where their favorite modality's performance was equal or better than any other. Throughout the case studies presented on this paper, the inclusion of time constraints made users try to optimize their interaction with the questionnaires. They started to use the most effective modalities whenever they could: direct interaction on the touch screen instead indirect interaction through their favorite modality; navigation between pages through their favorite modality; and, input on open ended answers through audio instead of text on the virtual keyboard.

On environments where the users were alone, or surrounded by few known persons, the type of content addressed by the questions (private or not) did not dictate the input modality used on open ended answers, in this situations our studies clearly point to a preference directed to audio recording instead of textual input through the virtual keyboard. However, the increasing of the people surrounding the users intimidated them, reducing or eliminating the usage of voice interaction with our application

(favorite interaction modality of the majority of the users involved in these studies). Initially, there was a reduction on the utilization of voice interaction when users, surrounded by few unknown persons, were answering private questions (as in the park's case study). After, when surrounded by more people (as in the subway's case study), users completely stopped using voice interaction because they were embarrassed of speaking to a device in front of so many people.

Even when visual output could not be used in optimal conditions (e.g. user walking or very intense light reflection on the screen), users decided not to use the audible counterpart of the artefact. The only situation where they decided to use it was while driving, because they could not spare any visual attention to the device.

Restrictions to hand usage, namely, the ones that force users to hold and interact with the device using the same hand (subway's case study, task 2) tend to introduce difficulties performing: direct interaction with the elements on the screen (mostly because the users' thumb fingers could not easily reach the whole area of the screen on our PDAs); and, most notably, textual input through a virtual QWERTY keyboard (because it is difficult to perform finger interaction on a virtual keyboard presented on the screen). In these situations users prefer to interact indirectly with the elements through their favorite modality and perform audio input on open ended questions (unless there are many persons around them).

Throughout the presented case studies there were few situations where users could not use textual input properly and didn't feel comfortable to answer certain questions through audio input (e.g. walking on the park or, most notably, standing on the metro). On these situations users could not find a combination of modalities through which they could complete their task comfortably. We believe that this problem could be solved through the introduction of a virtual T9 keypad, directed for one hand finger interaction on devices with no physical T9 keypad (as the one used on this study).

5 Conclusions and Future Work

In this paper we presented multimodal artefacts that allow user input through 5 different modalities. We report a study conducted in order to understand which interaction modalities are preferred by users in different contexts, considering different context variables. This study enabled us to understand some behavioral patterns defined by everyday contexts (such as home, park, subway and car) and context variables (lightening, movement, hand usage, etc.).

It is clear that some modalities suit better some users in some contexts, from this fact emerges the need of adapting mobile interaction and interfaces according to their utilization context. Our future work plans aim at the creation of intelligent context-aware multimodal adaptive artefacts.

References

1. Nicolle, C., Abascal, J. (eds.): *Inclusive Design Guidelines for HCI*. Taylor and Francis, London (2001)
2. Sá, M., Carriço, M.: *Defining Scenarios for Mobile Design and Evaluation*. In: *Procs. of CHI 2008 SIGCHI Conference on Human Factors in Computing Systems*, Florence, Italy. ACM Press, New York (2008)

3. Hurtig, T.: A mobile multimodal dialogue system for public transportation navigation evaluated. In: *Procs. of HCI 2006*, pp. 251–254. ACM Press, New York (2006)
4. Turk, M., Robertson, G.: Perceptual user interfaces introduction. *Commun. of the ACM* 43(3), 33–35 (2000)
5. Oviatt, S., Darrell, T., Flickner, M.: Multimodal interfaces that flex, adapt, and persist. *Commun. ACM* 47(1), 30–33 (2004)
6. Lai, J.: Facilitating Mobile Communication with Multimodal Access to Email Messages on a Cell Phone. In: *Procs. of CHI 2004*, pp. 1259–1262. ACM Press, New York (2004)
7. Blattner, M.M., Gliner, E.P.: Multimodal integration. *IEEE Multimedia* 14–24 (1996)
8. Signer, B., Norrie, M., Grossniklaus, M., Belotti, R., Decurtins, C., Weibel, N.: Paper Based Mobile Access to Databases. In: *Procs. of the ACM SIGMOD*, pp. 763–765 (2006)
9. Santoro, C., Paternò, F., Ricci, G., Leporini, B.: A Multimodal Museum Guide for All. In: *Mobile interaction with the Real World Workshop, Mobile HCI, Singapore* (2007)
10. Oviatt, S.: Mutual disambiguation of recognition errors in a multimodal architecture. In: *Procs. CHI 1999*, pp. 576–583. ACM Press, New York (1999)
11. Blenkhorn, P., Evans, D.G.: Using speech and touch to enable blind people to access schematic diagrams. *Journal of Network and Computer Applications* 21, 17–29 (1998)
12. Boyd, L.H., Boyd, W.L., Vanderheiden, G.C.: The Graphical User Interface: Crisis, Danger and Opportunity. *Q. Journal of Visual Impairment and Blindness* 84, 496–502 (1990)
13. Reis, T., Sá, M., Carriço, L.: Designing Mobile Multimodal Artefacts. In: *Procs. of 10th ICEIS, Barcelona, Spain*, pp. 75–89. INSTICC (2008)
14. Brewster, S.A.: Overcoming the lack of screen space on mobile computers. *Personal and Ubiquitous Computing* 6(3), 188–205 (2002)
15. Brewster, S.A., Lumsden, J., Bell, M., Hall, M., Tasker, S.: Multi-modal ‘eyes free’ interaction techniques for wearable devices. In: *Proc. CHI 2003 Conference on Human Factors in Computing Systems, CHI Letters*, vol. 5(1), pp. 473–480. ACM Press, New York (2003)
16. Palen, L., Salzman, M.: Beyond the Handset: Designing for Wireless Communications. *ACM Transactions on Computer-Human Interaction* 9, 125–151 (2002)
17. Makela, A., Giller, V., Tscheligi, M., Sefelin, R.: Joking, storytelling, art sharing, expressing affection: A field trial of how children and their social network communicate with digital images in leisure time. In: *Proceedings of CHI 2000*, pp. 548–555. ACM Press, New York (2000)
18. Kjeldskov, J., Skov, M.B., Als, B.S., Hoegh, R.T.: Is it Worth the Hassle? Exploring the Added Value of Evaluating the Usability of Context-Aware Mobile Systems in the Field. In: *Proceedings of Mobile HCI 2004, Berlin, Heidelberg*, pp. 61–73 (2004)
19. Beck, E.T., Christiansen, M.K., Kjeldskov, J.: Experimental Evaluation of Techniques for Usability testing of Mobile Systems in a Laboratory Setting. In: *Proceedings of OzCHI 2003, Brisbane Australia* (2003)
20. Goodman, J., Gray, P., Khammampad, K., Brewster, S.: Using Landmarks to Support Older People in Navigation. In: *Proceedings of Mobile HCI 2004, Verlag, Berlin, Heidelberg*, pp. 38–48 (2004)
21. Baillie, L., Schatz, R.: Exploring multimodality in the laboratory and the field. In: *Procs. of the 7th international conference on Multimodal interfaces, Toronto, Italy* (2005)

An Audio-Haptic Aesthetic Framework Influenced by Visual Theory

Angela Chang¹ and Conor O'Sullivan²

¹20 Ames St. Cambridge, MA 02139, USA
anjchang@media.mit.edu

²600 North US Highway 45, DS-175, Libertyville, IL 60048, USA
conor.o'sullivan@motorola.com

Abstract. Sound is touch at a distance. The vibration of pressure waves in the air creates sounds that our ears hear, at close range, these pressure waves may also be felt as vibration. This audio-haptic relationship has potential for enriching interaction in human-computer interfaces. How can interface designers manipulate attention using audio-haptic media? We propose a theoretical perceptual framework for design of audio-haptic media, influenced by aesthetic frameworks in visual theory and audio design. The aesthetic issues of the multimodal interplay between audio and haptic modalities are presented, with discussion based on anecdotes from multimedia artists. We use the aesthetic theory to develop four design mechanisms for transition between audio and haptic channels: *synchronization*, *temporal linearization*, *masking* and *synchresis*. An example composition using these mechanisms, and the multisensory design intent, is discussed by the designers.

Keywords: Audio-haptic, multimodal design, aesthetics, musical expressivity, mobile, interaction, synchronization, linearization, masking, synchresis.

1 Introduction

We live in a world rich with vibrotactile information. The air around us vibrates, seemingly imperceptibly, all the time. We rarely notice the wind moving against our bodies, the texture of clothes, the reverberation of space inside a church. When we sit around a conference table, our hands receive and transmit vibrations to emphasize what is being said or attract attention to the movements of other participants. These sensations are felt by our skin, a background symphony of subtle information that enriches our perception of the world around us.

In contrast, products like the LG Prada phone [22] and the Apple iPhone [1] provide little tactile feedback (figure 1). Users mainly interact with a large touchscreen, where tactile cues are minimal and buttons are relegated to the edges. This lack of tactile feedback causes errors in text entry and navigation [29]. In order to give more feedback, audio cues are often used to confirm tactile events and focus the user's attention [8], e.g. confirmation beeps. However, these audio cues are annoying and attract unwanted attention [16]. Many HCI researchers are now researching how haptics (physical and tactile) can provide a subtler feedback channel [5,13, 27, 28, 29].



Fig. 1. (a) LG Prada phone and (b) Apple iPhone are touchscreen based mobile devices

The use of haptics (particularly vibration) is promising, because it is relatively cost effective and easy to implement [5]. The benefits to a vibrotactile interface, mainly privacy and subtlety, are not new [3,5]. Yet, there is relatively little knowledge on how to aesthetically structure and compose vibrations for interfaces [4,14]. This work addresses the creation of multimodal experiences by detailing our experiences in developing haptic ringtones in mobile phones, through describing an example of audio-haptic stimuli. We share our knowledge and expertise on how to combine audio-haptic information to create pleasing and entertaining multimodal experiences.

2 Background and Motivation

Prior work in HCI has explored mapping vibration to information through the use of scientifically generated media [10,25]. Some work has focused on developing haptic hardware and identifying situations where tactile information can be used, e.g., navigating spatial data [12], multimodal art [15], browsing visual information, and of course, silent alerts [4,6,13,17]. We note the majority of creation techniques for vibrotactile stimuli have been largely designated by device capabilities. O'Modhrain[18] and Gunther [11] have presented works on composing vibrations in multimedia settings, using custom hardware. A key issue has been how to map information to vibration to avoid overload [17]. This work seeks to extend the prior art by suggesting a general aesthetic framework to guide audio-haptic composition.

One inspiration has been the Munsell color wheel [16]. The Munsell color wheel is a tool for understanding how visual effects can be combined. The use of the color wheel gives rise to color theory, and helps graphic designers understand how to create moods, draw attention, and attain aesthetic balance (and avoid overload). We wondered if a similar framework could help vibrotactile designers compose their audio-haptic effects so that there is stimulating, but not overwhelming transition between the audio and haptic modalities to create a complex, but unified multisensory experience.

Audio-visual theories for cinematography has also influenced this work, particularly, cinematic composer Michel Chion's theories about the relation of audio to vision [7]. *Synchronization* (when audio happens at the same time as visual event), *temporal linearization* (using audio to create a sense of time for visual effects), *masking* (using audio to hide or draw attention away from visual information) and the *synchresis* (use of sound and vision for suggesting or giving illusion to add value onto the moviegoing experience) aroused our interest. The idea behind audiovisual composition is *balance* and *understanding the differences between audio and visual through perceptual studies*.



Fig. 2. Two Audio-haptic displays using embedded (a) MFTs made by Citizen (CMS-16A-07), used in a (b) vibrotactile “puff”, and (c) the Motorola A1000

2.1 Audio-Haptic Media Design Process

The easiest approach to designing vibrotactile interfaces is to use low power pager motors and piezo buzzers. Multifunction transducers (MFTs) enable the development of mobile systems that convey an audio-haptic expressive range of vibration [20], figure 2a. An MFT-based system outputs vibrations with audio much like audio speakers.

Instead of focusing strictly on vibration frequencies (20Hz-300Hz) [26], we recognize that audio stimuli can take advantage of the overlap between vibration and audio (20Hz-20kHz), resulting in a continuum of sensation from haptic to audio called the *audio-haptic spectrum*. In the past, audio speakers were often embedded into the computer system, out of the accessible reach of the user. Mobile devices have allowed speakers to be handheld. By using MFTs instead of regular speakers, the whole spectrum of audio-haptics can be exploited for interactive feedback.

Two example devices were used in our process for exploring the audio-haptic space (figures 2b and 2c). One is a small squishy puff consisting of one MFT embedded in a circular sponge (2b). Another is the Motorola A1000 phone which uses two MFTs behind the touchscreen. Audio-haptic stimuli are automatically generated by playing sounds that contain haptic components (frequencies below 300Hz). If necessary, the haptic component could be amplified using haptic inheritance techniques. In our design process, we used commercially available sound libraries [21,24]. Audio-haptic compositions were made using Adobe Audition [2]. A user holding either the squishy puff or the A1000 phone would be able to feel the vibrations and hear audio at the same time.

2.2 Designing a Visual-Vibrotactile Framework

The Munsell color wheel (figure 3) describes three elements of visual design. One principle element of visual design is the dimension of “warm” or “cool” hues. Warm colors *draw more attention* than “cool colors”. In color theory, warm colors tend toward the red-orange scale, while cool colors tend towards the blue purplish scale. The warm colors are on the opposite side of the color wheel from cool colors. Another component of color theory is value, or the lightness amplitude in relation to neutral. The darker the color is, the lower its value. The final dimension is chroma, or the saturation of the color. The chroma is related to the radial distance from the center of the color wheel.

We wondered whether prior classifications of vibrotactile stimuli could provide a similar framework [4,14,19]. Scientific parameters such as frequency, duration and amplitude (e.g. 100 Hz sine wave for 0.5 seconds) have traditionally been used to describe vibrations in perception studies. Perceiving vibrations from scientifically

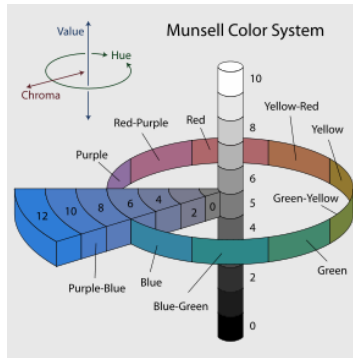


Fig. 3. Munsell Color System showing Hue, Value, and Chroma¹

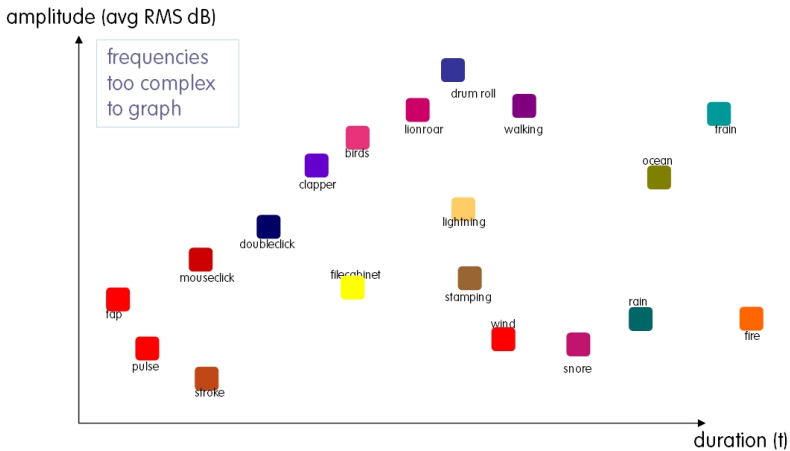


Fig. 4. Audio-haptic stimuli plot based on amplitude and vibration (scatter plot)

generated stimuli has some flaws, particularly since they are detached from everyday experience. These synthetic vibrations are unrelated to the experience of human interaction with objects. Users often have to overcome a novelty effect to learn the mappings. In contrast, normal everyday interactions with our environment and objects result in vibrations and sound. In this inquiry, we have elected to select stimuli based on sound stimuli from commercially available sound libraries [21, 24].

As a starting point, approximately 75 audio-haptic sounds were selected based on their audio-haptic experience. When laid out on a grid consisting of frequency, duration and amplitude, it was hard to organize these complex sounds based on frequency. Graphing the duration and amplitude produced a scatterplot, and did not suggest any aesthetic trends (figure 4 shows a plot with less sounds than our actual plot).

¹ Munsell Color System, <http://en.wikipedia.org/wiki/Munsellcolorsystem> on June 23, 2008.

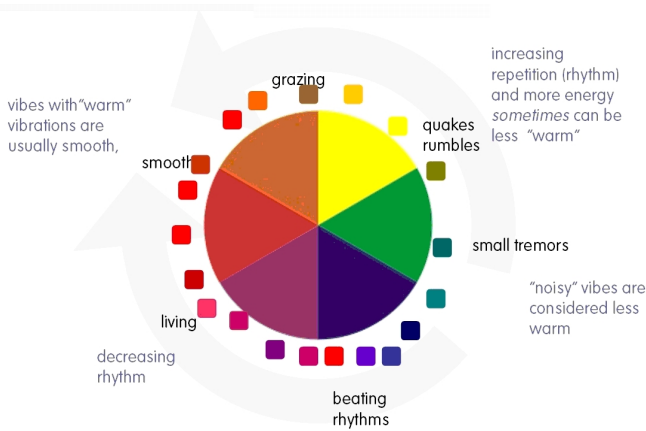


Fig. 5. Activity Classification for audio-haptics shows a number of problems: conflicting warmth trends, multiple classifications for similar sounds

Another prior framework for classifying audio-haptic media is activity classification [20], which characterizes audio-haptic media by context of the activity that may generate the stimuli. While the activity classification system is a good guide for users to describe the qualitative experience of the stimuli, it is hard for designers to use as a reference for composition. The main problem with this mapping was that the stimuli could belong to more than one category. For example, “fire” or “wind” sound could belong to both surface and living categories. The same stimuli could be considered complex or living. A new framework should contain dimensions that are quantitatively distinguishable. The stimuli were characterized according to the activity classification and graphed onto a color wheel to determine if there were any aesthetic trends, figure 5. There were conflicting trends for warmth or cool that could be discerned. Smooth sounds, such as pulses or pure tones could be considered warm, but “noisy” stimuli such as quakes or beating sounds could also be considered warm (drawing attention).

The Munsell color wheel suggests that energy and attention are perceptual dimensions for distinguishing stimuli. In the audio-haptic domain, a temporal-energy arrangement scheme was attempted by using ADSR envelopes. ADSR (Attack-decay-sustain-release) envelopes are a way to organize the stimuli based on energy and attention [23], figure 6. The attack angle describes how quickly a stimuli occurs, the decay measures how quickly the attack declines in amplitude, the sustain relates to how long the stimuli is sustained, and the angle of release can correspond to an angle of warmth of ambience, resulting in an “envelope” (figure 6a). Some typical ways to create interesting phenomena is to vary the ADSR envelope are shown (figure 6 b).

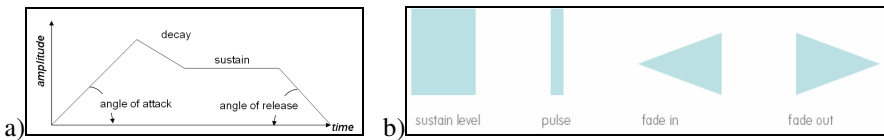


Fig. 6. a)ADSR envelope for audio design composition b) typical ADSR envelopes

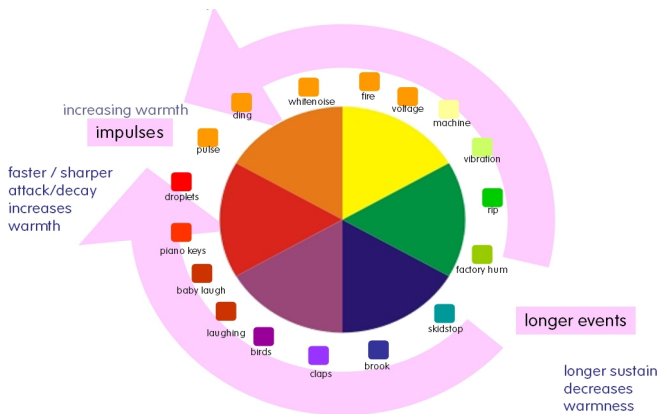


Fig. 7. The visual-vibrotactile audio-perception map temporal trend for the attack/decay and duration of a stimuli, increased sharpness and length attract attention.

We analyzed the average RMS value of the audio-haptic sounds, and gave a numerical value for the length of the stimuli [23]. A sound analysis tool, Adobe Audition [23], was used to analyze the envelope of each audio-haptic stimulus. Each stimulus was plotted atop the color wheel such that warmth (intensity) corresponded to attack/decay and amplitude. Higher intensity draws more user attention to the stimuli. A resulting visual-vibrotactile audio-perception map was generated (figure 7).

The correspondence between temporal-energy suggested an aesthetic trend based on amplitude and duration. The perception map was informally evaluated by 10 media artists using a squishy puff device connected to an on screen presentation. The artists were asked to select the mapped stimuli and evaluate the mapping out loud. General feedback on the visual-vibrotactile perception map was assessed to gauge how “natural or believable” the stimuli seemed.

Here are some aesthetic trends that were observed:

- Amplitude corresponds to saturation. The higher the amplitude, the more noticeable it is. In general, the amplitude effect dominates the stimuli.
- Attack/decay duration corresponds to intensity or haptic attention. Faster (sharper) attack/decays are more noticeable than smoother attack/decays. The smoother attacks are more subtle in haptic feel. However, there is a minimum length (10 milliseconds) where the skin cannot feel the vibration at all and the stimuli are perceived as pure audio and may be missed.
- Longer events, such as rhythmic or sustained stimuli, are also more noticeable (audibly “warmer”), but can be ignored over time as the user attenuates the stimuli. The skin can lose sensitivity to sustained sounds.

By noting the elements of composition as amplitude, attack/decay duration and sustain, **two main compositional effects** can be inferred from the visual and cinematic design literature:

1. *Balance*: Balancing interaction between high and low amplitude audio-haptic stimuli.
2. *Textural variation*: Alternating between impulse and sustained stimuli, and playing with the ADSR envelopes to create textural variation.

These two observed effects can be used as guidelines to create dramatic effects through manipulating attention between the two modalities.

3 An Audio-Haptic Composition Example

We present an audio-haptic ringtone to describe some composition mechanisms used to balance and textural variation, separately and concurrently (figure 8).

This piece explores the relationship between the musical ringtone and the vibrotactile alert. These mechanisms are introduced gradually, then developed - separately and overlaid - in the following short segments. Overall a sense of urgency is generated to fulfill the primary purpose of mobile alerting mechanism.

The piece begins with a haptic texture that has been created using the haptic inheritance method [5], where the haptic texture has been derived from the percussive instrumental elements. The texture can be classified as a type of pulse with a soft tail. The texture enters in a rhythmic manner, increasing somewhat in strength over the course of its solo 2 bar introduction. Then as the phrase repeats, the audible percussive element fades in a rhythmic unison.

When the percussive phrase is completely exposed (spectrally, dynamically and musically), a 2 bar answering phrase enters. This answering phrase contains no significant haptic texture but generates content in the broader frequency spectrum with the introduction of a musical sequence. The woody pulse that punctuates the phrase is “warm” in drawing attention to the rhythmic audio crescendo. Apart from the melodic and rhythmic phrase this sequence also contains a melodic decoration in the vein of a traditional auditory alert.

In the final section the haptic rhythm re-enters solo, apart from the appearance of the melodic alert decoration which is played at the same time in the rhythmic sequence as before.

The order of exposition is as follows:

1. Haptic Rhythm
2. Haptic Rhythm + Percussive Instruments
3. Percussive Instruments + Musical/Melodic Elements
4. Haptic Rhythm + Melodic Alert

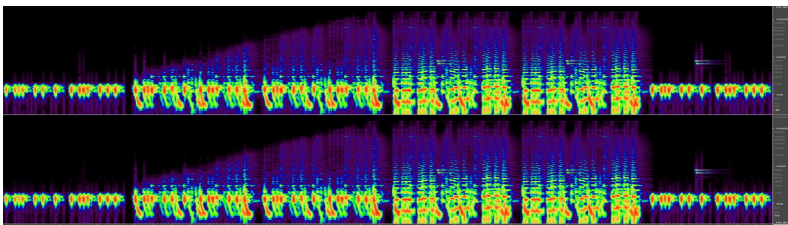


Fig. 8. Spectrogram of audio-haptic ringtone example

As such the envelope created is a type of privacy envelope. This can be inferred from the attached frequency spectrum of the piece; when the energy is concentrated in the lower range the privacy is greater, lessening over time, before a final private reminder is played.

4 Discussion

4.1 Aesthetic Perception Map Feedback

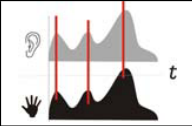
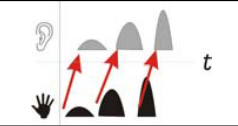
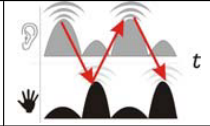
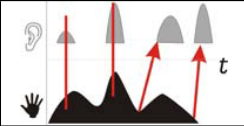
Overall, the feedback on our perception map was positive. Many artists commented that the effects of synchronizing haptics and audio were compelling, and that they could readily imagine the textures of the materials or ambience of the stimuli in the perception map. This suggests that *synchresis* (synchronizing and synthesizing) artificial audio-haptic stimuli could create realistic experiences that do not exist in the real world, a very interesting phenomenon [9]. Another common observation was that higher haptic amplitude present in the stimuli created a feeling of realism, particularly for impacts. The effect of audio reverb, combined with sustained and synchronized haptics seemed “more realistic” than without haptic feedback. For example, combined modal experiences allowed users to gauge the differences between metal objects hitting each other, vs. a hammer hitting an anvil. Another example is being able to distinguish a splashing “large rock in a stream” compared with a “drop of water hitting a large pool”. We believe the haptic component helped users imagine or identify the underlying material and environmental information in the stimuli.

One artist commented that perhaps amplitude was analogous to color saturation, drawing attention and perhaps even overwhelming the senses. She expressed some doubt that there could be three dimensions for audio-haptic perception. Rather, she suggested that instead of a color wheel, it could be a grayscale wheel that may be most appropriate. Another interesting issue is that visual warmth is achieved through reddish tones, which grab attention, but audio warmth is achieved through sustain duration, which fades as time progresses. Whether tactile warmth is affected by sustain duration is an interesting question. In general, feedback for the perceptual map was well perceived, but future work will need to verify its accuracy through multidimensional scaling.

4.2 Audio-Haptic Modality Transfer Discussion

Our example composition makes use of the four audio-haptic mechanisms between the dual modalities (Table 1). Here *temporal linearization* of the haptic elements create suspense, and *synchronization* with (percussion) audio allows transfer of the attention to the audio track. *Masking* is achieved by increasing the audio amplitude and spectral range over the haptic elements in the main body of the piece. *Synchresis* is created by the reverb of the audio elements, creating a sense of ambient resonance. We propose that distributing both warm and cool stimuli across the two modalities helps attain compositional *balance*. Similarly, *textural variation* is achieved by contrasting the sharpness of the synchronized syncopated percussion and haptic rhythm with the more sustained musical elements.

Table 1. Audio-haptic mechanisms used in achieving balance and textural variation

Synchronization	Temporal Linearization	Masking	Synchresis
Haptic and audio stimuli are united in <i>time</i> to draw attention to the unity of the stimuli.	Haptic or audio sequencing to create a connection between multisensory stimuli, to create a sense of causality. Usually haptic will precede audio.	The hiding of a (typically haptic) stimulus by another stimulus of higher amplitude when they are presented in together.	Creation using audio and tactile to create an association of an effect. Makes use of linearization and synchronization to create a mental union of two modalities, creating a distinct new association..
			

It should be noted that we are reporting the state of the art of how we practice audio-haptic design. We hope that this tutorial can be useful to help others struggling with audio-haptic composition. Discussion of the resulting artifact through use of the mechanisms was subjective to the ringtone designers, and has not been empirically tested. We look forward to contributing to a collaborative discussion on how to achieve audio-haptic balance and textural variation.

5 Conclusion

This paper presents the evaluation of audio-haptic stimuli from sound libraries as a starting point toward realizing the potential for audio-haptic composition. We discuss an approach to arranging the stimuli based on temporal-energy distributions, in a framework influenced by visual and cinematic design principles. Several perceptual features were observed by artists testing a perceptual map of audio-haptic stimuli. These observations resulted in the creation of guidelines for achieving balance and textural variation. These perceptual findings need to be supported by further studies using quantitative evaluation.

From the compositional guidelines we have developed four principle mechanisms for audio-haptic interaction design: *synchronization*, *temporal linearization*, *masking* and *synchresis*. We present an example of an audio-haptic ringtone composition that utilizes these mechanisms and discuss the multimodal effects achieved. We describe how these four approaches can be used to manipulate attention across modalities. Our example showcases a practical example of these mechanisms. In summary, we describe how compositions of carefully designed audio-haptics stimuli can convey a richer experience through use of the audio-haptic design mechanisms presented.

Acknowledgements

The authors thank Prof. John Maeda for the questions that started this research and Prof. Cynthia Breazeal for her advice. The authors would also like to express gratitude to Motorola's Consumer Experience Design group for their support of this work.

References

1. Apple iPhone, <http://www.apple.com/iphone/>
2. Adobe Audition 3, <http://www.adobe.com/products/audition/>
3. Brewster, S., Chohan, F., Brown, L.: Tactile feedback for mobile interactions. In: Proc. of the CHI 2007, pp. 159–162. ACM Press, New York (2007)
4. Brewster, S.A., Wright, P.C., Edwards, A.D.N.: Experimentally derived guidelines for the creation of earcons. In: Proc. of HCI 1995, Huddersfield, UK (1995)
5. Chang, A., O’Modhrain, M.S., Jacob, R.J., Gunther, E., Ishii, H.: ComTouch: a vibrotactile communication device. In: Proc. of DIS. 2002, pp. 312–320. ACM Press, New York (2002)
6. Chang, A., O’Sullivan, C.: Audio-Haptic Feedback in Mobile Phones. In: CHI 2005 Extended Abstracts, pp. 1264–1267. ACM Press, New York (2005)
7. Michel, C., Gorbman, C., Murch, W.: Audio-Vision Sound on Screen. Columbia University Press, New York (1994)
8. Gaver, W.: Chapter from Handbook of Human Computer Interaction. In: Helander, M.G., Landauer, T.K., Prabhu, P. (eds.) Auditory Interface. Elsevier Science, Amsterdam (1997)
9. Geldard, F., Sherrick, C.: The cutaneous rabbit: A perceptual illusion. *Science* 178, 178–179 (1972)
10. Geldard, F.A.: *Body English*. Random House (1967)
11. Gunther, E.: *Skinscape: A Tool for Composition in the Tactile Modality*. MIT MSEE Thesis (2001)
12. Jones, L., Nakamura, M., Lockyer, B.: Development of a tactile vest. In: Proc. of Haptic Symposium 2004, pp. 82–89. IEEE Press, Los Alamitos (2004)
13. Luk, J., Pasquero, J., Little, S., MacLean, K., Levesque, V., Hayward, V.: A role for haptics in mobile interaction. In: Proc. of CHI 2006, pp. 171–180. ACM Press, New York (2006)
14. MacLean, K.E.: *Designing with Haptic Feedback*. In: Proc. of IEEE Robotics and Automation (ICRA2000), San Francisco (2000)
15. Merrill, D., Raffle, H.: The sound of touch. In: CHI 2007. Extended Abstracts, pp. 2807–2812. ACM Press, New York (2007)
16. Munsell, A.H., Munsell, A.E.O.: *A Color Notation*. Munsell Color Co., Baltimore (1946)
17. Nelson, L., Bly, S., Sokoler, T.: Quiet calls: talking silently on mobile phones. In: Proceedings of CHI 2001, pp. 174–181. ACM Press, New York (2001)
18. O’Modhrain, S., Oakley, I.: Adding Interactivity. In: Proc. of the Int. Symp. on Haptic Interfaces for Virtual Environment and Teleoperator Sys. (HAPTICS 2004), pp. 293–294. IEEE Press, Los Alamitos (2004)
19. O’Sullivan, C., Chang, A.: An Activity Classification for Vibrotactile Phenomena. In: McGoekin, D., Brewster, S.A. (eds.) HAID 2006. LNCS, vol. 4129, pp. 145–156. Springer, Heidelberg (2006)
20. O’Sullivan, C., Chang, A.: Dimensional Design; Explorations of the Auditory and Haptic Correlate for the Mobile Device. In: Proc. of ICAD 2005, Limerick, Ireland, pp. 212–217 (July 2005)
21. Pinnacle Systems. Pinnacle Studio 10 Sound Library. [Mountain View, CA] (2005)
22. Prada phone by LG, <http://www.pradaphonebylg.com/>
23. Riley, R.: *Audio Editing with Cool Edit*. PC, Tonbridge (2002)
24. *Sound Library Sonothèque = Sonoteca*. [France]: Auvidis (1989)
25. Tan, H.Z.: Perceptual user interfaces: haptic interfaces. *Communications of the ACM* 43(3), 40–41 (2000)

26. Verrillo, R.T., Gescheider, G.A.: Tactile Aids for the Hearing Impaired. In: Summers, I. (ed.) *Perception via the Sense of Touch*, ch. 1. Whurr Publishers, London (1992)
27. Vogel, D., Baudisch, P.: Shift: a technique for operating pen-based interfaces using touch. In: *Proc. of the CHI 2007*, pp. 657–666. ACM Press, New York (2007)
28. Williamson, J., Murray-Smith, R., Hughes, S.: Shoogle: excitatory multimodal interaction on mobile devices. In: *Proc. of the CHI 2007*, pp. 121–124. ACM Press, New York (2007)
29. Zhao, S., Dragicevic, P., Chignell, M., Balakrishnan, R., Baudisch, P.: Earpod: eyes-free menu selection using touch input and reactive audio feedback. In: *Proc. of CHI 2007*, pp. 1395–1404. ACM Press, New York (2007)

In Search for an Integrated Design Basis for Audio and Haptics

Antti Pirhonen and Kai Tuuri

Department of Computer Science and Information Systems
FI-40014 University of Jyväskylä, Finland
{pianta, krtuuri}@jyu.fi

Abstract. Audio and haptics as interaction modalities share properties, which make them highly appropriate to be handled within a single conceptual framework. This paper outlines such framework, gaining ingredients from the literature concerning cross-modal integration and embodied cognition. The resulting framework is bound up with a concept of physical embodiment, which has been introduced within several scientific disciplines to reveal the role of bodily experience and the corresponding mental imagery as the core of meaning-creation. In addition to theoretical discussion, the contribution of the proposed approach in design is outlined.

Keywords: haptics, audio, integration, multimodal, embodiment.

1 Introduction

The personal computer owes a great deal to the typewriter in its design. The development of the contemporary variety of personal computers from the first generation PCs consists of a number of small steps - new features and devices have been added to the basic unit one after another. The process can therefore be characterised as evolutionary. The markets, in the first place, have defined whether a certain new feature has been viable or not. The important point here is that new features have been assessed in terms of their match with the prevailing practices. There have not been many opportunities to genuinely question the basic concept of PC, even though that device unnecessarily incorporates the limitations of an ancient technology.

The first PCs rapidly grew in capacity and gradually got new features which made them quite different from their ancestors. Once new technologies, such as audio, made it possible to present information in multiple ways, the talk about multimedia started and the related form of interaction became known as multimodal. Looking back at the development of multimodal user interfaces, it turns out that most of the research and development has been opportunistic and technology-driven by nature. Each time a new interaction technology has been introduced, the developers have soon found use for the new opportunities it can provide. In other words, interaction has been designed in terms of available technologies, rather than in terms of observed interaction needs.

The technology-centred approach has resulted in technical conceptualisation of multimodality and multimodal interaction. For instance, when a high-quality sound device was included in a PC, the PC was renamed a "multimedia workstation" whose use was argued to be multimodal by nature.

When multimodality is defined in technical terms, it is quite easy to find a connection to the early stages of human-computer interaction (HCI). In those days, HCI was mainly seen as a means to "synchronise" human being and a computer (e.g. Card et. al. [1]). While the number of computer users rose rapidly and computers were suddenly in the hands of ordinary people, there was an obvious need to make computers more easy to use than what they were when used by computer engineers. Psychologists were challenged to model human mind and behaviour for the needs of user-interface design. It was thought that if we knew how human mind works, user interfaces could be designed to be compatible with it.

In practice, the traditional HCI approach led to oversimplification of the conception of human mind. Until recently, the dominating metaphor of the human mind has been the computer [2]. Thus multimodality has often meant that in interaction with a computer, several senses ("input devices") and several motor systems ("output devices") are utilised. However, we argue that this kind of conceptualisation of human being as a smart device is extremely limited, and conflicts with both the contemporary view of human mind and the way we behave in our everyday environments. In this paper, we present various arguments to support our basic claims:

1. Interaction is always multimodal in nature.
2. Design arises from mental images and results in mental images.
3. Acknowledging the bodily nature of interaction as a cornerstone in design inevitably results in the support for multimodal interaction.

These claims define an approach to user-interface design. The proposed approach is based on a sound theoretical foundation, and is meant to be applicable by practitioners. In this paper we present the relevant background theories.

2 Modal Interconnections in Perception

The concept of multimodality and the related, more technical concept, multimedia, have been handled in literature in many ways. In this study, we are looking for an alternative approach to multimodality, focusing on the relationship between audio and haptics as interaction modalities. We start by figuring out the nature of multimodality, proceeding from the implications to design.

2.1 Conceptualisation of Multimodality

Multimodality can be conceptualised in terms of technical, cognitive, social or a number of other perspectives. In this sub-section, we briefly present typical examples as a background for an alternative view.

Multimodality as a Technical Opportunity. In the development of technical products, a typical procedure starts from new technical opportunities. Once the technology is there, we have to find uses for it. It can be argued that much of what is marketed as multimedia are products resulting from this kind of approach. Especially in the early stages of multimedia, the producers were under pressure to show their technical sophistication by supporting all available means of interaction.

Once a critical mass had been reached in the sales, multimedia products can be argued to have become part of our everyday life. The next step was to elaborate the multimedia technology. An essential part of the elaboration was to legitimate the technology in terms of human-computer interaction. Advantages were sought for from multimodal interaction. The models of human cognition on which multimodality conception was based were very simple. A typical example is an idea of a free cognitive resource; for instance when information was presented via a visual display, other sensory systems were thought of as free resources. When this claim was empirically found unsustainable, human ability to process information from multiple sources and in multiple modalities became a central issue. An important source of information was attention studies. In them, human ability to process information had been under intensive research since 1950's, when the rapidly growing air traffic made the cognitive capacity of air-traffic controllers the bottle-neck of fluent flight organisation. These studies resulted in models, which either modelled the structure of those mechanisms which define attention (structural models, the cornerstones found in [34]), or models which analysed human capacity (capacity models, see [5]).

It is worth noting that most of the applications of attention studies to multimodal interaction treat human being as a bottle-neck of a system. The studies of human capacity in this area can mostly be interpreted as studies of human limitations. We argue that this does not genuinely indicate human-centred approach.

Multimodality Provides Options. The use of multimedia or designing multimodal applications is often thought of as a selection of means of interaction. For instance, it is easy to find texts which give an impression that a given piece of information can be presented in various forms; text, speech, picture, animation etc. The underlying idea is that there is the content and there is the form that is independent of it. However, this notion has been found untenable in various disciplines. In the context of information presentation in user interfaces, it has been found that paralleling, e.g., sound and an image is extremely complicated. When trying to trace meaning creation on the basis of non-speech sound by asking the participants of an experiment to pair sounds and images, it was found that conclusions could be made only when the images were simple symbols indicating a clearly identifiable piece of information (like physical direction [67]). In other words, the idea that the designer is free to choose in which modality to present certain information is a grave oversimplification of the design process. Worn phrases like "the medium is the message" [8] or that "a picture is worth a thousand words", still hold in the multimodal context. Referring to the sub-heading, i.e., multimodality provides options, it should be understood as that

advanced technology provides options but that different options are qualitatively different. The process of choosing an interaction modality is not independent of other design efforts.

Multimodality Provides Redundancy. In mathematical information theory [9], redundancy was introduced as something to get rid of. Redundancy unnecessarily uses the resources of a communication channel, thus lowering the efficiency of an information system. However, in the context of information systems, the concept of redundancy has also more positive connotations; redundancy can be seen as a way of increasing system stability by providing backup.

This idea, which originates from the mathematical theory of communication, has been applied to human-computer interaction in the era of multimedia. It has been argued that if information is delivered in multiple formats, the message is more reliably received. A classic example is users with disabilities; if information is provided both in an audio and visual format, for example, the same application can be used as well by users with impairment in vision as by those with impairment in hearing [10].

As can be seen, the inseparability of form and content of information (discussed in the previous section) inevitably questions the endeavour of presenting "the same" information in multiple formats. Even if these kinds of redundant combinations undoubtedly are appropriate in many applications, they should not be seen as a straightforward, universal solution to the need of providing information in the right format to each particular user.

Natural Interaction is Multimodal. All the approaches discussed above are technically oriented in that they analyse modalities in terms of available technology. In a virtual world, the sound of a virtual object is created by a sound file which is processed with audio software run by audio hardware. The visual image of the same virtual object, in turn, is a product of an image file, a graphic processor and a visual display. In contrast, in the interaction with real life objects, such analysis is not appropriate - it is one single physical object which directly causes audio, visual, tactile and perhaps olfactory perception. Real world objects don't have separate devices to cause stimulus in different sensory modalities. Therefore, the interaction with real world objects is always multimodal by nature. This quite common sense notion has been used as a rationale for multimodal user-interfaces as well. However, the suggested naturalness cannot be achieved by burying the application under a heap of multimedia effects. The naturalness can only be achieved by designing objects which are not in the first place "sounds" or "images" or anything else which primarily refers to certain technology. Multimodality should not be an end in itself - rather, it should be treated as an inevitable way of interaction.

2.2 Embodied Meaning

Perspectives of phenomenology, pragmatism or ecological perception have considered meaning as being based on our interactions with the world rather than as

a separate abstract entity. Embodied perspective continues that line of reasoning and rejects the traditional Cartesian body-mind separation altogether: "terms body and mind are simply convenient shorthand ways of identifying aspects of ongoing organism-environment interactions" [11]. Cognition is thus seen as arising inherently from organic processes. Imagination, meaning, and knowledge are structured by our constant encounter and interaction with the world via our bodies and brain [12,13]. The perspective of embodied cognition seeks to reveal the role of bodily experience as the core of meaning-creation, i.e., how the body is involved in our thinking.

From the embodied point of view, the central aspect of meaning is action. Thus the relationship between action and perception is considered as close. So-called motor theories of perception (see e.g. [14]) has suggested that we understand, e.g., what we hear, because we somehow senso-motorically resonate the corresponding action by imaging the way the sound is produced. Several contemporary studies (see Gallese and Lakoff [13] for a review) in neuroscience indeed suggest that perception is coupled with action on a corporeal basis. Discoveries of common neural structures for motor movements and sensory perception has elevated the once speculative approach into a more plausible and appealing hypothesis. According to mentioned studies, all sensory modalities are integrated not only with each other but also together with motor control and control of purposeful actions. As a result, doing something (e.g. grasping or seeing someone grasping) and imaging doing it activates the same parts of the brain.

Gallese and Lakoff [13] argue that to be able to understand something, one must be able to imagine it, i.e., mentally simulate corresponding action(s). In other words, understanding is action-orientated mental imagery, and meaning thus equals the way something is understood in its context. Other authors have also suggested [15,16] that understanding involves a mental re-enactment or simulation of what we perceive. According to studies reviewed by Gallese and Lakoff [13] the action simulation seems to occur in relation to 1) motor programmes for successful interaction with objects in locations, 2) intrinsic physical features of objects and motor programmes (i.e., manners) to act on them to achieve goals (i.e., general purposes), and 3) actions and intentions of others. The "mirrored" simulation of motor movements of other people is interestingly hypothesized to act as a mechanism for empathy [17], i.e., perception of mental and motivational states.

As we can see, the perspective of embodied cognition provides convenient insights at least for the creation of concrete and affective action-related meanings and concepts. However, it is also suggested [18] that even concepts of language could be based on the stable patterns of embodied multimodal sensory-motor experiences. Such higher level preverbal foundations of linguistic meanings are referred to as image schemas.

2.3 Embodied Multimodality and Mental Images

The arguments of embodied cognition clearly have an effect on the conception of multimodality that was discussed in 2.1. The understanding of an action (like

grasping) is multimodal in the sense that action is neurally enacted using shared neural substrate for perception and action, which also responds to more than one sensory modality. Such view of multimodality denies the existence of separate modules - i.e. separate sensory inputs and motor control which do not integrate until in the presumed higher "associating area". Multimodality thus seems to be a fundamental core property of our perception and thinking where the linkage of performance and perception as well as the integration of sensory modalities is a cognitive norm.

From the viewpoint of mediated communication the multimodal nature of understanding means that similar imaginary experience of "grasping" could be triggered by using either visual, haptic or auditory cues with a suitable action-specific affordance. Separate presentation modalities can either function co-operatively (with integrated action relevance) or by themselves. Regardless of the presentation mode, physical articulation with various hints of, e.g., movement, direction, force, object properties in the presentation can make us to imagine actions that in some way make sense in the current situational context.

The embodied view of multimodality and understanding explains the cross-modal associations based on one presentation modality - for example, why music can create imagery of patterns of movement, body gestures, force and touch [19,20]. In a similar way, visual presentation can also evoke, for example, haptic meanings. However, maybe the most prominent contribution of the embodied view is bringing interaction, the experience of bodily encounters with its physical constraints and invariants, motor cognition and goal-orientation into the core of action perception.

On the whole, the action-orientated perspective of embodied cognition is in the same line with the current development in the philosophical foundation of HCI-field emphasising everyday (ecological/social) experiences, situated actions and the ideas of tangible interaction [21]. It also has a great importance in regard to interaction design, suggesting that acting/doing performs an equal part along the perception in the meaning creation.

3 Designing Action-Relevant UI-Elements by Utilising Mental Imagery

On the theoretical basis presented in this paper, we argue that designing sound or haptics for human-computer interaction involves the communication of action-relevant mental imagery. This imagery is multimodal human experience, and as such it should be considered as the starting point in UI-element design by exploring and specifying the relevant mental imagery in relation with communicational purposes, events and processes in interaction. Hence, we argue that the starting point should not be rigidly any specific presentation modality but the interaction and holistic exploration of the embodied "imagery" of its meanings. These meanings rise from purposeful actions of the user as well as from observed feedback effects and other actions presented by the user interface. However, we are not only concerned of concrete perceptions of action, but also of imaginary experiences that

become coupled with those actions. Of course, the UI-designer cannot define these couplings for sure, but consistent interaction-centric scenarios of imagery in her mind would form a design basis for articulation via UI-elements, resulting in likely communication of relevant imagery. This perspective is closely related to the usage of metaphors in design (which is very different from the common usage of the term metaphor in the context of graphical user-interfaces, see [22]).

To achieve the intended action-relevance, it would be wise to consider the relationship between perceived/imaginary action and auditory/haptic presentation. Jensenius [23] has suggested a distinction between natural action-sound couplings and artificially formed action-sound relationships. Natural coupling can be understood as plausible mechanical affordances between a sound and related action with involved objects. Moreover, Jensenius [23] suggests that "mental images of (natural) action-sound couplings guide our perception of artificial action-sound relationships". This is important, because convenient usage of mental images can act as a familiar mediator between abstract system processes and their contextual presentation to the user, and can provide the needed stability between often artificial relationships of actions and UI-presentation in HCI.

Godøy [24] has proposed a model of musical imagery where our understanding of (sound-producing) actions is founded on sensory-motor images of

- *excitation* (imagery about what we do or imagine/mimic doing - e.g., body movement or gestural articulation) and
- *resonance* (imagery of space, objects and materials about the effects of what we do or imagine/mimic doing).

Because of the multimodal, holistic nature of mental images, this model of musical imagery should easily be adaptable into a wider perspective that includes sound and haptics. The model can be used as a guide to explore both motor and material/object-related images of the involved activity. The aim of such exploration is ultimately to get ideas of suitable auditory and/or haptic cues that match with the action imagery and provide affordances to it. We can presume that such cues of action-oriented non-linguistic form of information can be transcoded relatively easily from one presentation modality to another. Or be composed in co-operative integration of both modes.

To complement the model presented above, we suggest an additional form of imagery to be considered. That is the imagery of intention, which concerns goals, motivations and (why not) also emotions behind the perceived/imaginary activity. The exploration of intention should be closely linked to the designer's meta-design considerations of functional purposes of each instance of auditory/haptic element, and to how this instance is intended to be "framed" into the actual situational use of an application. Indeed, keeping interaction design in mind, we even suggest this functional meaning of the UI-element to be the first and foremost factor to be considered. Generally, the first set of questions to be asked is: What should be the perceived function (general purpose) of the auditory/haptic instance? And to what situational indexes and especially to what user actions the sound or haptic instance is associated? The functional dependency to the situational context may vary. With a suitable situational index, even minimal

action-relevant parameters (e.g. force or direction) of an auditory/haptic expression should facilitate the interaction.

In the case of feedback UI-elements, the central focus concerning intention can be focused to the motivations and plans behind user actions, because the UI-element should support the user's mental idea of doing and achieving something. In some cases, it may also be fruitful to explore intentions from the perspective of an imaginary "agent", i.e., the counterpart of interaction. For example, gestural imagery and articulation can be utilized as a basis for UI-elements that are meant to persuade the user to do something or inform the user about something (see [25]). A similar perspective can be applied also in specifying gestural recognition for the UI-input elements.

To sum up this design process and utilisation of mental images on a general level, we decided to categorise the different design phases by conforming to the model of creative process. Traditionally the process is described in five phases: 1) preparation, where the problem or goal is acknowledged and studied, 2) incubation, where the ideas are unconsciously processed, 3) insight, where the idea of a plausible solution emerges, 4) evaluation, where the idea is somehow concretised and exposed to criticism and 5) elaboration, where the idea is refined and implemented [26]. Our modified process includes four phases: 1) *defining the communicative function*, which pretty well matches with the classic preparation phase, 2) *exploration of action-relevant mental images*, which comprises recursive incubation and insight phases, 3) *articulation*, which refers to concretisation and evaluation of ideas, and finally 4) *implementation of media element*, which is the phase of the actual UI-element production.

4 Concluding Statements

Some years ago we designed a portable music player, which was supposed to be used without resorting to gaze. Its gaze free interaction was implemented with simple gestures and feedback sounds. Browsing of playlist, play/stop and volume controls consisted of taps and simple sweeps across the touch screen of a PDA device. The directions (forward-backward) were illustrated with panned feedback sounds. E.g., to select the next track, the user made a sweep forward on the touch screen resulting in a sound with increasing pitch moving from left to right part of the headphones. The sweeps were directed physically forward and backward, because the device was hanging on the side of the user in the seam of a pocket.

This experiment revealed important issues about the conceptualisation of physical directions and the related metaphors. In the experiment, we suddenly switched the feedback sounds, but only one of the ten participants even noticed it. We were wondering how it is possible that the chosen audio metaphors were such weak cues. In the very early stage we noticed that the directional metaphors of gestures and audio were in conflict with each other; the gestures were in the forward-backward direction, while the sounds relied on left-right direction as is the case with typical music players' control panel. We concluded that the reason

for rejecting the left-right metaphor might be that the gestures require physical activity, thus making the relating metaphor dominate over the conflicting audio metaphor.

From the point-of-view of embodied cognition, our observation deserves a fresh look. The notion of the central role of bodily experience in human perception and action indeed confirms our speculation about the reason of the dominance of gestures over audio feedback. This should not be interpreted as a finding for gestural interaction against other modalities, though. Rather, we see the contribution of our finding in stressing the central role of physical experience in all interaction design. As illustrated in above, we argue that an essential early phase of design is to articulate the design idea (or mental image or metaphor) as a physical action. This kind of description would then work as a sound basis for design, which naturally takes into account the verified bodily basis of our cognitive system. The proposed approach also questions some ideas about multimodality and the underlying computational metaphor of human cognition with its separate input and output systems. These observations seem to indicate that e.g. the models of attentional capacity and redundant information presentation should not be rejected but they could be understood in a new way in the framework of embodied cognition. Embodied cognition as an approach to design has thus potential in shifting the orientation of traditional, computation based models to something which could be called human-centred.

Acknowledgments. This work is funded by Finnish Funding Agency for Technology and Innovation (www.tekes.fi), and the following partners: Nokia Ltd., GE Healthcare Finland Ltd., Sunit Ltd., Suunto Ltd., and Tampere city council.

References

1. Card, S.K., Moran, T.P., Newell, A.: *The Psychology of Human-Computer Interaction*. Lawrence Erlbaum Associates, Hillsdale (1983)
2. Gardner, H.G.: *The Mind's New Science*. Howard Gardner, New York (1985)
3. Broadbent, D.E.: *Perception and Communication*. Pergamon, London (1958)
4. Deutsch, J.A., Deutsch, D.: Attention: Some Theoretical Considerations. *Psychological Review* 70(1), 80–90 (1963)
5. Wickens, C.D.: Processing Resources in Attention. In: Parasuraman, R., Davies, D.R. (eds.) *Varieties of Attention*, pp. 63–102. Academic Press, Orlando, F (1984)
6. Pirhonen, A.: Semantics of Sounds and Images - Can They Be Paralleled? In: *Proc. of International Conference on Auditory Display Schulich School of Music*, pp. 319–325. McGill University Montreal, Canada (2007)
7. Pirhonen, A., Palomäki, H.: Sonification of Directional and Emotional Content: Description of Design Challenges. In: *Proc. of International Conference on Auditory Display, IRCAM, Paris* (2008)
8. McLuhan, M.: *Understanding Media: the Extensions of Man*. New American Library, New York (1966)
9. Shannon, C.E., Weaver, W.: *The Mathematical Theory of Communication*. The University of Illinois Press, Urbana (1949)

10. Edwards, A.D.N.: Redundancy and Adaptability. In: Edwards, A.D.N., Holland, S. (eds.) *Multimedia Interface Design in Education*. NATO ASI Series F: Computer and System Sciences, vol. 76, pp. 145–155. Springer, Heidelberg (1992)
11. Johnson, M., Rohrer, T.: We Are Live Creatures: Embodiment, American Pragmatism, and the Cognitive Organism. In: Zlatev, J., Ziemke, T., Frank, R., Dirven, R. (eds.) *Body, Language, and Mind*, vol. 1, pp. 17–54. Mouton de Gruyter, Berlin (2007)
12. Lakoff, G., Johnson, M.: *Philosophy in the Flesh: The Embodied Mind and its Challenge to Western Thought*. Basic Books, New York (1999)
13. Gallese, V., Lakoff, G.: The Brain's Concepts: The Role of the Sensory-Motor System in Reason and Language. *Cognitive Neuropsychology* 22, 455–479 (2005)
14. Liberman, A.M., Mattingly, I.G.: The Motor Theory of Speech Perception Revised. *Cognition* 21, 136 (1985)
15. Wilson, M., Knoblich, G.: The Case for Motor Involvement in Perceiving Conspecifics. *Psychological Bulletin* 1(3), 460–473 (2005)
16. Godoy, R.I.: Motor-Mimetic Music Cognition. *Leonardo* 36(4), 317–319 (2003)
17. Gallese, V.: Embodied Simulation: From Mirror Neuron Systems to Interpersonal Relations. In: *Empathy and Fairness (Novartis Foundation Symposium)*, vol. 278, pp. 3–19. Wiley, Chichester (2006)
18. Johnson, M.: *The Body in the Mind: The Bodily Basis of Meaning, Imagination, and Reason*. University of Chicago, Chicago (1987)
19. Godøy, R.I.: Gestural Imagery in the Service of Musical Imagery. In: Camurri, A., Volpe, G. (eds.) *GW 2003. LNCS (LNAI)*, vol. 2915, pp. 55–62. Springer, Heidelberg (2004)
20. Tagg, P.: Towards a Sign Typology of Music. In: Dalmonte, R., Baroni, M. (eds.) *Secondo Convegno Europeo di Analisi Musicale*, pp. 369–378. Università Degli Studi di Trento (1992)
21. Dourish, P.: *Where the Action is: The Foundations of Embodied Interaction*. MIT Press, Cambridge (2001)
22. Pirhonen, A.: To Simulate or to Stimulate? In Search of the Power of Metaphor in Design. In: Pirhonen, A., Isomäki, H., Roast, C., Saariluoma, P. (eds.) *Future Interaction Design*, pp. 105–123. Springer, London (2005)
23. Jensenius, A.: *Action-Sound: Developing Methods and Tools to Study Music-Related Body Movement*. Ph.D. Thesis. Department of Musicology, University of Oslo (2007)
24. Godøy, R. I. Imagined action, excitation, and resonance. In: Godøy, R. I. and Jørgensen, H. (eds.): *Musical Imagery*. Swets and Zeitlinger, Lisse (2001) 237-250
25. Tuuri, K., Eerola, T.: Could Function-Specific Prosodic Cues Be Used as a Basis for Non-Speech User Interface Sound Design. In: *Proc. of International Conference on Auditory Display, IRCAM, Paris* (2008)
26. Csikszentmihalyi, M.: *Creativity: Flow and the psychology of discovery and invention*. HarperCollins, New York (1996)

tacTiles for Ambient Intelligence and Interactive Sonification

Thomas Hermann^{1,2} and Risto Kõiva²

¹Ambient Intelligence Group, CITEC, Bielefeld University

²Neuroinformatics Group, Faculty of Technology, Bielefeld University
Postfach 10 01 31, D-33501 Bielefeld, Germany
{thermann,rkoiva}@techfak.uni-bielefeld.de

Abstract. In this paper we introduce *tacTiles*, a novel wireless modular tactile sensitive surface element attached to a deformable textile, designed as a lay-on for surfaces such as chairs, sofas, floor or other furniture. *tacTiles* can be used as interface for human-computer interaction or ambient information systems. We give a full account on the hardware and show applications that demonstrate real-time sonification for process monitoring and biofeedback. Finally we sketch ideas for using *tacTiles* paired with sonification for interaction games.

1 Introduction

Cognitive Interaction Technology (CIT) reshapes our concepts of human-computer interaction, going away from the paradigm of explicit control of technical products towards a seamless closure of interaction loops between a human user and an intelligent system that cooperatively engage in activity. We observe a trend of increasing intelligence in devices, ranging from wearables to smart rooms that surround the users.

Generally, such systems need input channels to sense the environment including the user, and output channels in order to communicate to the user. In most current technical systems, keyboard, mouse or touchpad interaction is used as input and visual displays are used as output (think for instance of mobile phones, DVD players, computers, etc.) so that technology couples primarily via eyes & hand. The interactions presented in this paper make use of different channels, namely ears & body, using the registration of tactile patterns and responding by using *sonification*, the auditory display of information.

In this paper we present *tacTiles* as a new device to equip surfaces with tactile sensitivity, allowing the computer to experience continuous spatially resolved pressure profiles in contact with a human user. Similar to how our interaction profits from tactile sensitive hand and body surface, *tacTiles* allow to extend the surface of CIT-artefacts such as intelligent rooms or furniture, making them tactile sensitive and allowing to use the information to better understand a system's environment or its use. Our first *tacTiles* prototype has been developed as a lay-on for chairs (Fig. [1](#)), however, a connection with other everyday-objects such as sofas, mattresses, doormats, yoga mats, backpacks etc. are equally intended.



Fig. 1. tacTiles as lay-on for office chairs

Tactile sensor-equipped chairs have already been considered before, such as the SenseChair [1], the sensingChair [2,3], and references can be found to tactile input mats. We see the innovation of tacTiles mainly in the following regards: firstly, tacTiles is an open-source available cheap solution, wireless and versatile to be used in many varying contexts without modifications. Secondly, tacTiles can be combined to form larger patches, e.g. to cover larger sofas, the floor, walls, etc. Thirdly, we here explore the potential of interactive sonification as feedback channel, for instance to induce a specific interaction pattern.

Applications for tacTiles depend on the use context: as a lay-on on chairs, they can be used to monitor via an ambient sonification the ergonomics of behavior (as presented in Sec. 6), to identify the user via its total pressure (corresponding to weight), to enable novel sorts of security checks (e.g. password entrance via a motion shift sequence on the seat). As a floor mat in front of a standard computer workplace, it allows to use the feet to interact with the computer - think for instance of a painting program where you can interactively modify the brush size by foot pressure. As a door mat, it allows to register who enters or leaves a room, e.g. as alarm system or support system for room intelligence to shut off the heat in unused rooms.

tacTiles as floor mats allow novel types of interactive games which are now also explored by companies such as Nintendo with the Wii Fit¹; however, the connection of tactile sensing and interactive sonification presented here is new and allows *eyes-free audiomotoric games* as introduced by the authors within the scope of *AcouMotion*, a system to explore similar audio-motion couplings [4]. tacTiles can even be used to couple several users by registering their interactions with the mat and reflecting for instance differential information as sonification

¹ <http://www.nintendo.com/wiifit>, last seen 2008-04-27

so that a synchronous performance can be practised. Such applications could even be interesting for training programs in dance or performance.

The design and applications presented in this paper therefore represent only a proof-of-principle into the field of tactile-audio-mediated interactions, and we focus here on the application of tacTiles as lay-on for chairs. The hard-/software of tacTiles is presented in Sec. 2, and some interaction tests are presented in Sec. 3. We proceed in Sec. 4 with a short explanation of interactive sonification as the basis of auditory display for closing the interaction-loop. In Sec. 5 we show how dimension reduction of interaction patterns can reveal typical use patterns, suitable as controls for subsequent sonifications. We then show some exemplary applications of tacTiles. The paper closes with a discussion of our approach and an outlook on future work.

2 TacTiles

This section describes the hardware of tacTiles, including the design ideas for the sensor mat and the electronic circuits to acquire and process measurements into a bluetooth-broadcasted sensor data stream. We plan to distribute the design of tacTiles as open-source hardware, providing circuits and assembly instructions online at www.sonification.de/publications/HermannKoiva2008-TFA by the date of the HAID conference. Finally we show how sensor data is parsed by receiver software to obtain the real-time data stream for storage, processing and sonification.

2.1 tacTiles Sensor Mat Design

Our prototype tacTiles sensor, a lay-on for office chairs, incorporates 8 force sensitive resistors (FSR) from IEE² of type CP-154, each utilising 40x40 mm active sensing area, with saturation point of 100 N/cm² (Fig. 2 a).

The positions of the sensors were extracted from tests with numerous persons of different height and size to optimize the area of greatest contact with the seat and back area. As derived from the experiments, 4 sensors were placed on the seating area in X pattern, 4 on the back with T pattern for measuring the degree of leaningness of a user (Fig. 2 b). The CP-154 sensors are glued to a 5 mm thick foam (material also used for producing camping mats), which forms a sturdy base protecting the sensors from over-elongation. The foam is cut into the form to fit typical office chairs seating and back area. The 2 flexible wires from each FSR-sensor run on the backside of the mat, glued with Pattex Express into the carved grooves, to the signal processing electronics, located on the backside of the lay-on seat (Fig. 2 c). The grooves in the foam protect the cabling and at the same time allow the user not to sense them. On top layer of the mat, where the sensors are located, additionally a thin layer of silk is glued with numerous glue spots around the sensors to avoid shearing and tearing the sensors from their positions. Rubber coated textile, sewn in a form of a bag, with side zipper for easy access to the interior is used to enwrap the foam with glued sensors.

² <http://www.iee.lu/>, last seen 2008-04-27

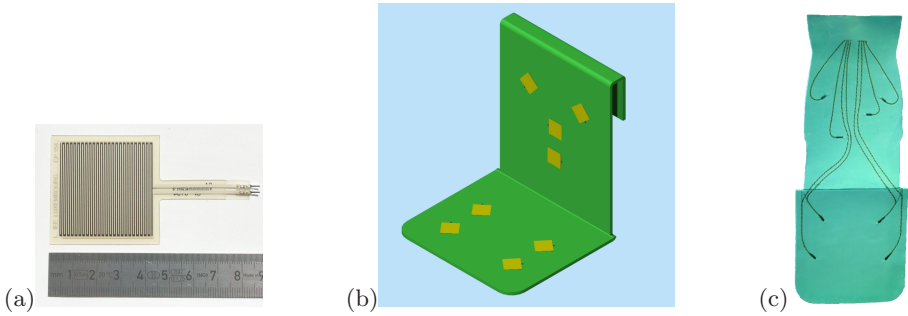


Fig. 2. Illustration of the mat sensors and mat assembly, here for the mat used as lay-on for office chairs. (a) CP-154 Force Sensitive Resistor used in tacTiles, (b) CAD model of the mat showing sensor placements, (c) Photo of the foam mat showing the routing of cables between sensor and electronics on the backside of the mat (right).

2.2 tacTiles Electronics Board

The electronics unit located behind the backrest of the chair (Fig. 3) consists of a Microchip PIC18F4580 microcontroller with an integrated 10-bit A/D converter, where the 8 FSR's values are read with the help of a pull-up resistor-network to produce voltage output of the sensors. The converted values are sent as an ASCII stream to Free2Move³ F2M01 Bluetooth Serial-Adapter, which is connected to the enhanced universal synchronous receiver transmitter (EUSART) output pin (TX) of the microcontroller. Thanks to the F2M01 module, the tacTiles have wireless coverage of more than 10m indoors, even through walls, easy interfacing thanks to Bluetooth Serial Port Profile (BT-SPP) supported by numerous operating systems, including many PDA and Smartphone models. Optionally the module allows encrypted communications, important for instance for motion-based authentication applications. The user interface of the electronics is derived from KISS (Keep-It-Simple&Stupid) ideology, incorporating just ON/OFF buttons, status LEDs and a charging jack.

Our tacTiles prototype is powered by 2-Cell Lithium-Ion Polymer battery with 700 mAh capacity, providing power for up to 8h of continuous operation. The microcontroller monitors the battery voltage, warns the user of low battery status and shuts the unit off, if the voltage gets critically low, important to prolongue the lifespan of the LiPo battery.

The readout values of 8 FSR sensors plus the battery voltage are sent using the ASCII protocol shown in Fig. 4. ASCII characters A to H are used to identify tactile sensors (A-D for seat, E-H for back), the battery voltage is identified with Z. The designators are followed by 3 ASCII numbers from 000 to 999, where higher values correspond to higher force applied onto the sensors. The battery voltage value is decimated after the first digit (673 means 6.73 V). The protocol is ended with line feed (LF) and carriage return (CR) for the possibility to have

³ <http://www.free2move.se/>



Fig. 3. Electronics unit located on the backrest of the chair

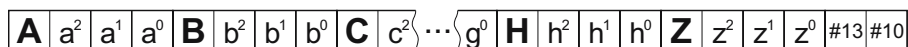


Fig. 4. Communications protocol in ASCII. A-H are tactile sensor identifiers, Z designates battery voltage.

easier look at the values with any terminal program. At a combined sampling rate of 10Hz for all inputs, the electronics utilizes very low data rate (3800 baud/sec.) and at the same time still achieving a crisp performance with no noticeable lag in reaction. The theoretical maximum sampling rate with the used electronic components and 8 sensors is 300 Hz.

3 Interaction Tests and tacTile Monitoring

In this section we demonstrate the operation of tacTiles at hand of a basic monitoring application. The presented applications are developed in Neo/NST, a cross-platform visual programming and rapid prototyping environment by Ritter et al. [5]. Fig. 5 depicts the GUI for the basic controller application that shows the sensor data as pressure profile and histogram plot (left/middle) together with computed dipole vectors for back and seat area (right). The application runs in real-time and allows furthermore to append recorded data vectors to a matrix that can be stored for later processing. In the video examples shown in the following section it can be seen that the latency is low enough to ensure real-time operation.

4 Interactive Sonification

Sonification is the scientific method of representing data by using non-speech sound (see [6] for a definition). Sonifications enable the listener to interpret structures in data from the auditory patterns that correspond to patterns in

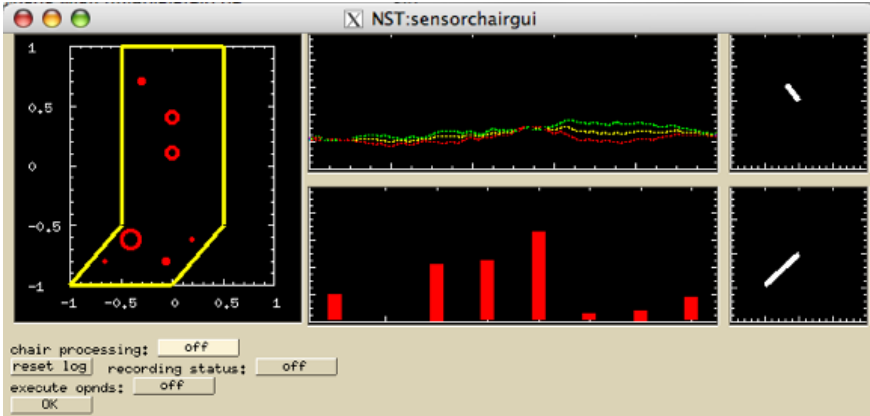


Fig. 5. Screenshot of the real-time tacTiles Controller application showing real-time pressure profiles (left plot), sensor sums for (back, seat, total) as (red, yellow, green) time series in the middle, and back and seat dipole vectors on the right side.

data. For instance, rhythmical patterns or changes therein are easier detected by listening than by visual inspection. Furthermore, listening allows eyes-free use which is beneficial in a variety of applications, ranging from information systems for visually impaired users to assistance systems for users (e.g. surgeons) whose visual focus is already highly occupied.

Interactive Sonification represents a focus subfield in sonification where the emphasis is set on how directly the interaction loop is closed between a system and the human user [7]. For instance, if any human action leads immediately to an acoustic change of the sonification, the user can better use the feedback to refine his/her activity. Applications for interactive sonification are widespread, e.g. training systems in sports [8], interactive control of physics experiments [9] or novel sports games using the AcouMotion system [4].

The most common technique to transform a real-time data stream into sound is by means of parameter mapping sonification, which uses actual sensor readings to drive a set of acoustic parameters of a sound synthesizer. Parameters can for instance be frequencies, amplitude, vibrato, pulse rate, etc. of sound events. We demonstrate a sonification for monitoring operation of the tacTile in Sec. 6.1.

5 Data Mining on tacTile Input Patterns

With only 8 sensors as in our prototype, it is still feasible to manually adjust a mapping. However, for higher-resolution tacTiles, it is useful to plug-in data mining techniques to reduce the dimensionality of the sensor data stream to a lower number of more meaningful features. We utilize principal component analysis (PCA) for that purpose, leading to the definition of eigenpatterns (or in case of a lay-on on chairs: eigenseats) that allow to decompose an actual pressure profile into a superposition of uncorrelated patterns. Fig. 6 shows a plot

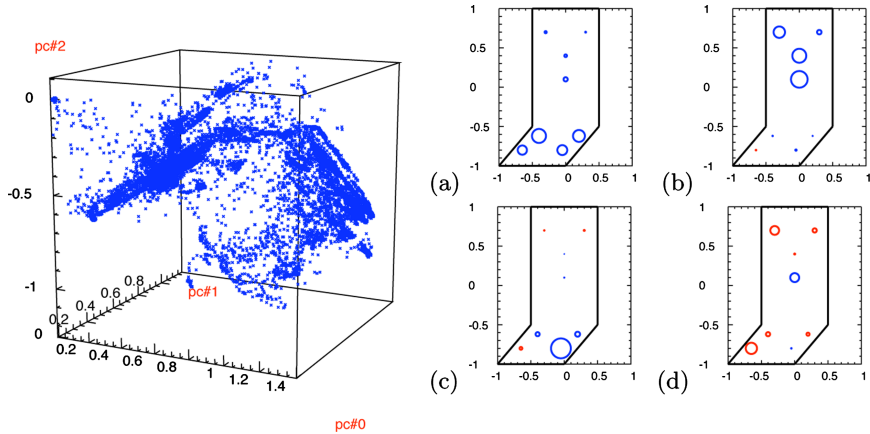


Fig. 6. PCA-plot of a 1h session on a tacTile equipped office chair. Obviously there are clusters corresponding to different activities such as typing or leaning back to relax. (a) shows the mean activity, (b,c,d) the first 3 eigenvectors, coding negative/positive components in red/blue.

of the first 3 PCA components of a data set from a 1h working session on the tacTiles equipped office chair (activity: the author writing this paper), and from the plot it is obvious that there are clusters of activity. Mapping these PCA-driven features to sound can facilitate the auditory identification of different characteristic states (clusters) by listening to the sonifications, compared to a mapping of raw sensor readings to sound stream parameters.

6 tacTile Applications

This section gives some practical examples for applications that are quickly realized by using tacTiles in different contexts. All applications use sonification and thus an auditory display channel. The sonifications are computed in real-time by using SuperCollider3, examples and videos of the interactions are provided on the website⁴.

6.1 Sonification for Real-Time Monitoring of Working Styles

As a first application we demonstrate tacTiles as a lay-on for an office chair to provide a real-time sonification of working style. As most straightforward parameter mapping sonification, we use 8 parallel audio streams corresponding to the 8 sensors. Each audio stream is a pulsed sequence of pitched tones in stereo space where pulse rate, pitch, amplitude, brilliance, stereo panning and pulse length can be controlled. We use pitch and panning to facilitate the identification of channels, mapping the sensor y -position (along the longer dimension) to pitch

⁴ Website: <http://www.sonification.de/publications/HermannKoiva2008-TFA>

so that higher located sensors on the chair give a higher pitched pulse sequence, and mapping the x -position (left-right on the chair) to the corresponding location in stereo space. This is understood intuitively and easy to learn. A continuously playing sound track would become annoying very quickly. For this reason, we use an *activity-driven* sonification, using the absolute derivative of sensor readings as input to a leaky integrator whose sum is mapped to the amplitude of some oscillators. In result, only sensor changes become audible and the sonification becomes automatically silent in case the tacTiles data remains constant. The pressure itself is mapped to the pulse rate so that sensors that experience a higher pressure are heard as pulsing at a higher rate, similar to a Geiger counter that ticks more frequently on higher radioactivity level. This sonification leaves variables such as the brightness or pulse duration unused, leaving headroom for future task-specific refinements that for instance might increase the brilliance with variance over a time-window, etc.

Interaction example S1⁵ shows the use of the monitoring sonification. Obviously, the sonification makes sense and reflects the movements on the chair. There is little immediate benefit from receiving such a direct auditory feedback on the sitting style, however, the controls are useful to check proper operation of the chair, and perhaps, if all chairs would be equipped like that in a large office space, a sum sonification could give a nice ambient impression of work patterns. Perhaps such sonifications might be interesting for visually impaired users to gain a sense of activity of visitors/interlocutors.

6.2 Rapid Scanning of Working Styles

This sonification represents recorded data such as a whole working day on an office chair equipped with tacTiles in a short time of a few seconds. This allows to review and summarize the interaction and work pattern very quickly. In the sonification example S2 we compress one hour of recorded tacTiles data into a sonification of 10 seconds, allowing to perceive what overall interaction pattern the subject has shown. Used here merely as a demonstration of the utility of sonification to rapidly summarize large sensor data series, it may prove useful for instance to rapidly scan sleep behavior for sleep monitoring stations where the patients' behavior could easily be recorded by using tacTiles as sensor mattress.

6.3 Emo-Feedback: Stay Tuned with EmoChair

In this sonification, basically an extended period of unchanging pressure profile is being interpreted as unhealthy, portrayed by an ambient information display involving sound and visuals in form of a facial feedback for the estimated degree of 'inflexibility'. Every physical activity on the chair is used to recharge a sensors' activity accumulator. If they fall under a threshold, increasingly motivating sound events start and the facial expression turns into a sad emotion. Being then more active on the chair rapidly recharges the happiness and stops the accumulator sonification. Fig. 7 shows a screenshot of visuals for unbalanced sitting for an extended

⁵ On our website.

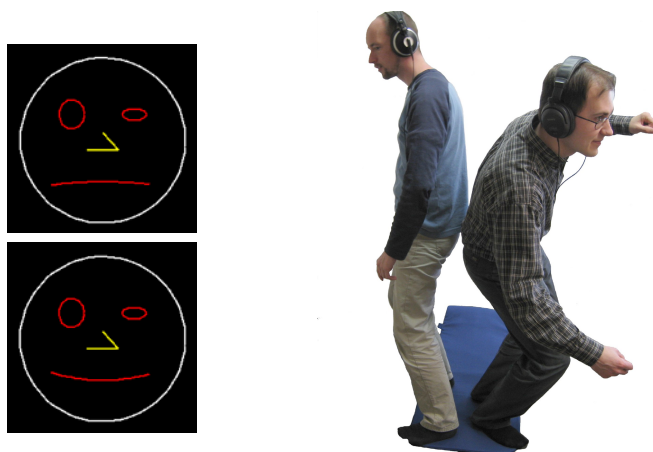


Fig. 7. tacTiles in application: the emo-faces (left) communicate bad mood if the user remains too rigid. The photo on the right side shows an interaction with tacTiles during the sonification-induced balancing-game.

period, and balanced, active sitting with more average activity over a time window. We present a video (S3 on the website) where the gradual change of sound and visuals with interaction is demonstrated. Experiments to verify in how far such ambient information displays may help to improve the average behavior, or beyond that, the long-term fitness of users, will be considered in the future after we have developed both the sensing and the sonification a bit further.

6.4 Sonification-Based Cooperative Balancing

Finally, we demonstrate the use of our tacTiles prototype without modification for an auditory cooperative balancing game. Two users rest with their feet on the tacTiles sensor mat, facing in opposite directions. Our first game idea is to keep a simulated boat in balance on a rough sea. The boat angle is displayed auditorily using sonification. Both players need to cooperatively balance their body weight in order to compensate for the external disturbance to keep the boat balanced. Since their own activity influences the balance of the boat, they can achieve a balancing using sound-induced body motion. The goal is to keep the boat in balance as good as possible. As a more competitive variation we plan a game where one user can try to shake the boat and the other player is challenged to stabilize it as fast as possible by corrective movements. Again, the interaction loop between the players is solely closed via interactive real-time sonification.

Games like that can help to train or develop body balance, or they finally become a sportive entertainment for both visually impaired and sighted players. Their eyes-free nature makes them ideal not only for visually impaired players, but also in settings where visual displays are unsuitable.

7 Discussion and Conclusion

In this paper we have presented tacTiles as a versatile, wireless, flexible sensor mat to be used as sensory skin for everyday-objects such as a sofas, mattresses, doormats, yoga mats, backpacks, etc. We have described the hardware, a control interface allowing real-time sonification, a sonification-based ambient interface to induce more dynamic working style on office chairs, a rapid scanning sonification, and finally, sketched an idea towards using tacTiles as interaction platform for sound induced motion games. The applications are in an early state yet the examples are already a proof-of principle. tacTiles allow a straightforward, wireless, cheap extension of passive rooms to more sensitive, ambient intelligent smart rooms.

As interface to register the user's activity our tacTiles as lay-on for chairs offer the potential to enhance the sensory data of our existing Augmented-Reality-based recording system⁶ to study alignment in human-human cooperation: the dynamic pressure profiles of interacting users may reveal synchronizations of behavior or bodily attention signals which contribute to the phenomenon of alignment.

Beyond potential uses of tacTiles as sensorial interface, *sonification* offers an ambient information display that does not bind the users' attention to a specific display location. We regard the combination of such audio-tacTiles as promising for future applications ranging from the medical field over entertainment to applications for visually impaired users.

Acknowledgements

We thank Mrs. Höppner for sewing the outer bag of the tacTile sensor mat. We thank Carsten Schürmann for soldering the PCB. We thank Oliver Lieske for his initial help with ideas and sensor placement tests on the mat. We thank Helge Ritter for providing the very stimulating work environment that formed the basis for the original implementation.

References

1. Forlizzi, J., DiSalvo, C., Zimmerman, J., Mutlu, B., Hurst, A.: The sensechair: the lounge chair as an intelligent assistive device for elders. In: DUX 2005: Proceedings of the 2005 conference on Designing for User eXperience, vol. 31, AIGA: American Institute of Graphic Arts, New York (2005)
2. Hong, Z., Tan, L.A.S., Pentland, A.: A sensing chair using pressure distribution sensors. IEEE/ASME Transactions on Mechatronics 6(3), 261–268 (2001)
3. Anttonen, J., Surakka, V.: Emotions and heart rate while sitting on a chair. In: Proc SIGCHI conference on Human factors in computing systems, CHI 2005, Portland, Oregon, USA (2005)

⁶ s. <http://www.sfb673.org>, project C5.

4. Hermann, T., Höner, O., Ritter, H.: Acoumotion - an interactive sonification system for acoustic motion control. In: Gibet, S., Courty, N., Kamp, J.-F. (eds.) GW 2005. LNCS (LNAI), vol. 3881, pp. 312–323. Springer, Heidelberg (2006)
5. Ritter, H.: The graphical simulation toolkit Neo/NST (2000), http://www.techfak.uni-bielefeld.de/ags/ni/projects/neo/neo_e.html
6. Hermann, T.: Taxonomy and definitions for sonification and auditory display. In: Katz, B. (ed.) Proc. Int. Conf. Auditory Display (ICAD 2008), France (2008)
7. Hermann, T., Hunt, A.: An introduction to interactive sonification (guest editors' introduction). IEEE MultiMedia 12(2), 20–24 (2005)
8. Höner, O., Hermann, T., Grunow, C.: Sonification of group behavior for analysis and training of sports tactics. In: Hermann, T., Hunt, A. (eds.) Proceedings of the International Workshop on Interactive Sonification (ISon 2004), Bielefeld, Germany, Bielefeld University, Interactive Sonification Community peer-reviewed article (2004)
9. Martini, J., Hermann, T., Anselmetti, D., Ritter, H.: Interactive sonification for exploring single molecule properties with afm-based force spectroscopy. In: Hermann, T., Hunt, A. (eds.) Proceedings of the International Workshop on Interactive Sonification (ISon 2004), Bielefeld, Germany, Bielefeld University, Interactive Sonification Community peer-reviewed article (2004)

An Audio-Haptic Interface Concept Based on Depth Information

Delphine Devallez¹, Davide Rocchesso², and Federico Fontana¹

¹University of Verona, Department of Computer Science
Strada Le Grazie 15, 37134 Verona, Italy
delphine.devallez@univr.it, federico.fontana@univr.it

²IUAV, Department of Art and Industrial Design
Dorsoduro 2206, 30123 Venezia, Italy
roc@iuav.it

Abstract. We present an interaction tool based on rendering distance cues for ordering sound sources in depth. The user interface consists of a linear position tactile sensor made by conductive material. The touch position is mapped onto the listening position on a rectangular virtual membrane, modeled by a bidimensional Digital Waveguide Mesh and providing distance cues. Spatialization of sound sources in depth allows a hierarchical display of multiple audio streams, as in auditory menus. Besides, the similar geometries of the haptic interface and the virtual auditory environment allow a direct mapping between the touch position and the listening position, providing an intuitive and continuous interaction tool for auditory navigation.

Keywords: Audio-haptic interface, auditory navigation, distance perception, spatialization, digital waveguide mesh, virtual environment.

1 Introduction

1.1 The Use of Distance Information

While most research dedicated to the design of new auditory interfaces focuses on directional spatialization of multiple sound sources [1,2,3,4,5], we propose an interface based on distance information. The motivation is driven by the ability of depth information to provide a hierarchical relationship between objects and therefore bring the attention of the user on the closest sound source. We believe that auditory interfaces would take great advantage of the depth dimension to provide additional information on the spatial layout of sound sources and therefore improve the organization of the auditory scene. In 1990, Ludwig [6] already suggested that techniques used in the music industry, such as reverberation and echo, could be valuable to the ordering of multiple sound sources in auditory interfaces.

1.2 Auditory Distance Perception

Auditory distance cues include intensity, direct-to-reverberant energy ratio, spectrum and binaural cues (Refer to [7,8] for detailed reviews of distance cues). Their

respective contributions to distance perception may differ according to the nature of and the familiarity with the sound source and the environment, as well as the availability of other non-acoustical cues. In general, distance perception is much less accurate than directional localization. Nevertheless, we are rather interested in the perception of the relative distances between multiple sound sources, assuming that the user is able to discriminate the respective positions of the sound sources. The studies of Strybel and Perrott [9] and Zahorik [10] touch upon the human perception of distance changes by measuring the resolution of source distance with the intensity cue and the direct-to-reverberant energy ratio cue respectively. Both studies suggested the possibility that the aforementioned cues represent distance changes, although the threshold of the direct-to-reverberant energy ratio is much coarser than the threshold of the intensity cue.

1.3 Related Work

With the ability to manipulate the spatial relationships between sound sources in depth, users may be able to focus their attention on a specific sound source corresponding to the closest distance to him or her. This manipulation is closely related to the technique called *acoustic zooming*, which accomplishes the focus on a specific object out of a multitude of sound sources. This is generally done by increasing the intensity of an object, for example as a function of its relative distance to the pointer controlled by the user [11,12], or as a function of the direction of the source compared to the user's median plane [3,5].

Instead of manipulating directly distance cues such as the intensity, which may also be manipulated by the user and therefore make this cue unreliable, spatialization techniques may offer better alternatives for distance rendering. In particular, recent studies have shown the ability of physical modeling of acoustic propagation to simulate acoustical environments. Of interest is the study conducted by Fontana and Rocchesso [13] which has underlined the effectiveness of a Digital Waveguide Mesh (DWM) modeling a rectangular parallelepiped to provide acoustic depth cues. Listening experiments using this model have shown its ability to render the apparent distance of sound sources. The only drawback of the proposed model was the high amount of computational resources required to simulate the environment, which did not allow a realtime application.

1.4 Scope of the Paper

The goal of this work is to explore a novel interaction tool whose main properties are the auditory feedback based on distance information provided by a digital waveguide mesh, and the direct mapping between the tactile user input and the auditory virtual environment. This paper focuses on the concept of the tool by describing the system, while the experimental assessment will be ulteriorly published. The haptic user interface is presented in Sect. 2 and provides the input listening position to the auditory spatialization process, described in Sect. 3. Possible applications of the resulting interaction tool are considered in Sect. 4.

2 The Haptic User Interface

The user interface consists of a ribbon controller, the Infusion Systems *SlideLong*, inspired by music controllers. This linear position tactile sensor has an active area of $384 \times 20 \text{ mm}^2$ and gives a value depending on the position at which the touch is made. A gamepad plays the role of sensor interface and is connected to the USB-port of the computer. As underlined by Jensenius et. al [14], such a game controller has the advantages of being cheap and having analog inputs which comply to the 0-5 volt sensor outputs. Besides connecting the sensor outputs on the motherboard is easy and the device uses the Human Interface Device driver supported in Max/MSP, which allows to make a fast, simple and low-cost interface sensor out of the game controller. Figure 1 shows the setup.



Fig. 1. Picture of the setup

The incoming values in Max/MSP are read by the built-in *human interface* object and are scaled to float numbers between 0 and 256. Since the ribbon has a rectangular geometry, it allows an easy analogy with the geometry of the rectangular DWM simulating the auditory environment. Therefore, after scaling, the incoming value from the ribbon controller directly provides the listening position input to the computation of the auditory signals in the DWM. In this way, a coherent mapping is performed between the touch position on the ribbon and the position of a virtual microphone in the DWM, and by moving the finger on the ribbon the user may explore the virtual environment where different audio streams are being attributed different positions. Like music controllers, this touch sensor intends to provide an interface that is intuitive to use with immediate and coherent response to user's gesture.

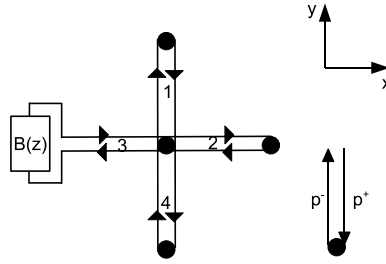


Fig. 2. Zoom on a node which is connected to other nodes via waveguides 1, 2 and 4. Waveguide 3 leads to a partially absorbing boundary. Triangles filled in black represent oriented unit delays.

3 Modeling of the Audio Space

3.1 The Acoustic Environment

Our proposed virtual acoustic environment consists of a rectilinear two-dimensional mesh whose digital waveguides simulate acoustic wave transmission between each internal junction. Each waveguide models the wave decomposition of a pressure signal p into its wave components p^+ and p^- , and each lossless node scatters 4 input signals coming from orthogonal directions, p_1^+, \dots, p_4^+ into corresponding output signals p_1^-, \dots, p_4^- (see Fig. 2). The properties of the wall materials contribute to the acoustics of a 3D space. This is also the case for a bidimensional acoustic environment since horizontal waves interact with the surface boundaries. Reflections from the boundaries are modeled by Digital Waveguide Filters (DWF), whose coefficients have been tuned to model specific reflective properties of real surfaces [15]. Finally, the number of nodes can be converted into the corresponding membrane dimensions once the speed of sound and the sampling frequency of the simulation have been determined. The model has been implemented in Max/MSP as an external object for realtime simulations.

3.2 Acoustical Properties of the Membrane

In order to investigate the auditory distance cues inside the virtual environment, the mesh dimensions are chosen to be 81×5 nodes, which correspond to $89.1 \times 4.4 \text{ cm}^2$, and impulse responses are computed at different distances on the membrane. The sound source is assumed to be point-wise and is located at the second node near one of the mesh widths. Measurements of impulse responses are carried out at diverse nodes on the y-axis of the membrane. Refer to Fig. 3 for the source and measurement positions.

Simulations of the listening environment was carried out in Matlab. Figure 4 shows the frequency responses up to 5 kHz, measured respectively at 12.1 cm and 72.6 cm.

Overall Intensity Level. Figure 5 shows the variation of the total energy with distance. By comparing with the energy decrease in open space, characterized by a

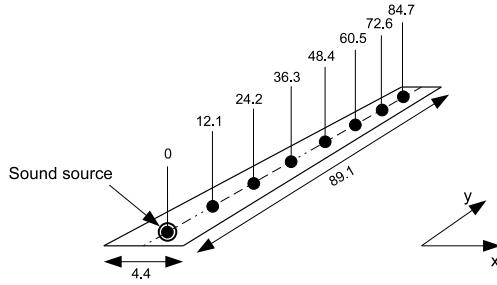


Fig. 3. The virtual membrane showing the measurement distances from the sound source position. All sizes are in centimeters.

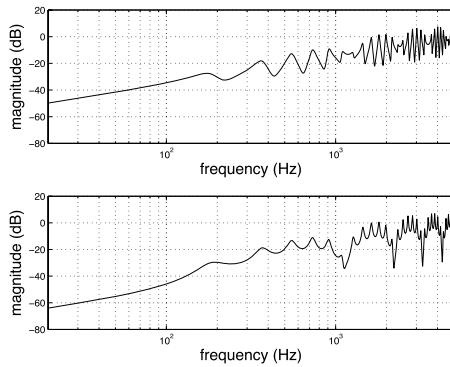


Fig. 4. Frequency responses up to 5 kHz on the membrane. Top: 12.1 cm. Bottom: 72.6 cm.

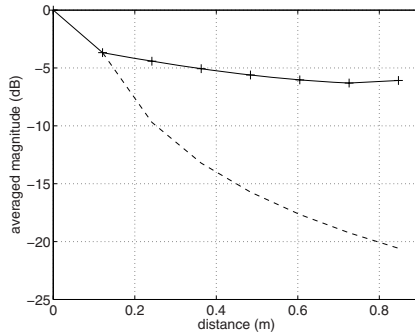


Fig. 5. Average magnitude of the impulse response as a function of distance. Solid line: 2D mesh. Dashed line: Reference open space.

reduction of 6 dB per distance doubling [16], it can be seen that the overall intensity on the membrane decreases significantly less. This behavior allows to hear even the farthest sound sources at any location on the membrane. In addition, the volume

can be manipulated by users, which might make the intensity cue unreliable for judging distance. Another reason for limiting the intensity cue for distance judgment is that the level of direct sound varies both with distance and with the energy emitted from the sound source, so that the listener needs some a priori knowledge about the sound source level in order to evaluate its egocentric distance.

Direct-to-Reverberant Energy Ratio. The virtual acoustic space we have designed aims at rendering depth information mainly thanks to the direct-to-reverberant energy ratio cue. For each impulse response, the delay of the direct sound is deduced from the distance between the sound source and the listening point, and is then removed from the impulse response. Afterwards the direct energy is integrated among the first 2.5 ms of the delay-free impulse responses, which approximates the duration of Head Related Impulse Responses measured in anechoic conditions and therefore captures the direct path of the sound signal [17]. Finally, the reverberant energy is calculated from the tail of the delay-free impulse responses. Figure 6 shows the values of the direct-to-reverberant energy ratios for different distances on the mesh. For comparison the direct-to-reverberant energy ratio v was computed for a natural environment, modeled by Zahorik with the function $v = -3.64 \log_2(r) + 10.76$ [10]. The two curves follow the same trajectory, suggesting that the direct-to-reverberant energy ratio in the virtual environment follows a natural behavior. Moreover, the values of the ratios are much lower in the 2D mesh than in the natural auditory space, which means that the amount of reverberation is exaggerated in the virtual environment.

The other known auditory distance cues, namely the Interaural Level Difference (ILD) and the spectrum, are not provided by the present model. This is motivated by their relatively weak contribution to distance perception. First ILDs, which arise due to intensity differences between the two ears, are null on the median plane and therefore will not provide any distance information for sound sources directly in front or behind the listener [18] as it is the case in our virtual environment. About the spectrum changes due to the air attenuation of high frequencies, they occur for very large distances (superior to 15 m) which lie beyond the available length of the mesh. As a result, the only pieces of information

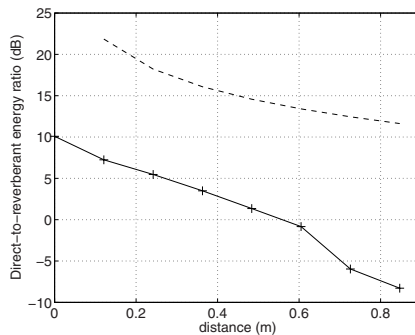


Fig. 6. Direct-to-reverberant energy ratio as a function of distance. Solid Line: 2D mesh. Dashed line: Approximation of a natural environment.

about the distance of a sound source in the mesh are the intensity and mostly the direct-to-reverberant energy ratio.

4 Applications

The audio-haptic interface presented in this paper allows navigation among multiple sound sources based on distance information. The continuous interaction is performed by a direct mapping between the position of the user's finger on the ribbon and the position of a virtual microphone in the mesh. In particular a hierarchical ordering of the audio data set is provided by the relative distance information of each audio item.

An auditory linear menu may originate from spatialization of auditory menu items in the DWM associated to carefully designed sonifications of the menu items, such as earcons, auditory icons, speech or spearcons. Without any spatialization layer it was shown that multiple concurrently presented earcons jeopardize the identification of the menu items [19]. On the contrary, using the human ability to localize sound sources in space may offer benefits to multi-stream rendering for identification tasks. In their study, Pitt and Edwards [12] showed that a linear menu with 2 to 8 sounds presented simultaneously and panned between the left and right channel produced a fastest average search time compared to the condition where only the sound associated to the pointer's position was played at a time. They also suggested that the task was more natural with concurrently presented sound sources, similar to identification of sound sources in real life situations.

Obviously the proposed auditory interface may also be used for audio browsing. The *Sonic Browser* developed by Fernström and Brazil [5] is a representative example of such an audio application. The spatial location of sound sources is rendered by stereo panning, i.e. loudness difference between the left and right channel. Further assistance to sound localization is given by a function that defines the range of perception, called *aura*. More precisely, it allows to get a finer discrimination between adjacent sound sources, acting like a zoom, in addition the radius of the aura is controllable by the user. The evaluation of the prototype [20] showed that simultaneous rendering of multiple stereo-spatialized sound sources allowed to complete the task faster than single-stream rendering. This result gives support to the use of our proposed auditory interface for audio browsing, where the length of the DWM would correspond to the size of the aura. In this way, the mesh would play the role of an acoustic lens rendering a few sound sources as a function of their distance to the user's finger on the ribbon, while another slider could be used for moving the acoustic lens along the whole set of audio files. The existence of two levels of resolution was already proposed [11] as a technique to avoid cacophony by limiting the number of concurrent sound sources and as a consequence speed up the completion of the user's task.

5 Conclusion and Future Work

We have proposed a tool for designing auditory interfaces based on spatialization of audio data in depth. The virtual environment rendering distance cues of sound sources is modeled by a rectilinear Digital Waveguide Mesh and has been implemented as an external in Max/MSP. Associated with a rectangular ribbon playing the role of a tactile input to the interface, this coherent tool may offer new opportunities for designing auditory interfaces. Future work consists in developing applications using this audio-haptic interface concept. In particular, a user study of audio browsing is planned in the near future.

Among the numerous issues raised by the tool presented in this paper, the nature and the number of sound sources to be displayed in the mesh need to be properly chosen. In addition to hardware and software restrictions about the length of the DWM, a compromise must be found between the confusion that may arise from too many sound sources rendered simultaneously, and the navigation efficiency.

Acknowledgments. We would like to thank our colleagues Stefano Papetti and Stefano Delle Monache for their valuable help. This research work has been supported by the European project FP6-NEST-29085 CLOSED - Closing the Loop of Sound Evaluation and Design.

References

1. Savidis, A., Stephanidis, C., Korte, A., Crispian, K.: A Generic Direct-Manipulation 3D-Auditory Environment for Hierarchical Navigation in Non-Visual Interaction. In: Proc. of the Second annual ACM conference on Assistive technologies, pp. 117–123. ACM, New York (1996)
2. Walker, A., Brewster, S., McGookin, D., Ng, A.: Diary in the sky: A spatial audio display for a mobile calendar. In: Proc. of the IHM-HCI, pp. 531–539. Springer, Heidelberg (2001)
3. Schmandt, C.: Audio hallway: A virtual acoustic environment for browsing. In: Proc. of the 11th annual ACM symposium on User interface software and technology, pp. 163–170. ACM, New York (1998)
4. Brewster, S., Lumsden, J., Bell, M., Hall, M., Tasker, S.: Multimodal ‘Eyes-Free’ Interaction Techniques for Wearable Devices. In: Proc. of the SIGCHI conference on Human factors in computing systems, pp. 473–480. ACM, New York (2003)
5. Fernström, M., Brazil, E.: Sonic Browsing: An Auditory Tool for Multimedia Asset Management. In: Proc. of the 2001 International Conference on Auditory Display, Laboratory of Acoustics and Audio Signal Processing and Telecommunications Software and Multimedia Laboratory, Helsinki University of Technology, Espoo, Finland, pp. 132–135 (2001)
6. Ludwig, L.F.: Extending the Notion of a Window System to Audio. *Computer*, 66–72 (1990)
7. Zahorik, P.: Auditory Display of Sound Source Distance. In: Proc. of the 2002 International Conference on Auditory Display. Advanced Telecommunications Research Institute (ATR), Kyoto (2002)

8. Nielsen, S.H.: Distance Perception in Hearing. PhD thesis, Aalborg University, Denmark (1991)
9. Strybel, T.Z., Perrott, D.R.: Discrimination of relative distance in the auditory modality: The success and failure of the loudness discrimination hypothesis. *J. Acoust. Soc. Am.* 76(1), 318–320 (1984)
10. Zahorik, P.: Direct-to-reverberant energy ratio sensitivity. *J. Acoust. Soc. Am.* 112(5), 2110–2117 (2002)
11. Winberg, F., Hellström, S.O.: Designing Accessible Auditory Drag and Drop. In: Proc. of the 2003 conference on Universal usability, pp. 152–153. ACM, New York (2003)
12. Pitt, I.J., Edwards, A.D.N.: Pointing in an Auditory Interface for Blind Users. In: Proc. of the 1995 IEEE International Conference on Systems, Man and Cybernetics, pp. 280–285. IEEE Press, New York (1995)
13. Fontana, F., Rocchesso, D.: A Physics-based Approach to the Presentation of Acoustic Depth. In: Proc. of the 2003 International Conference on Auditory Display, pp.79–82 (2003)
14. Jensenius, A.R., Koehly, R., Wanderley, M.M.: Building Low-Cost Music Controllers. In: Kronland-Martinet, R., Voinier, T., Ystad, S. (eds.) CMMR 2005. LNCS, vol. 3902, pp. 123–129. Springer, Heidelberg (2006)
15. Fontana, F.: Physics-based models for the acoustic representation of space in virtual environments. PhD thesis, University of Verona, Italy (2002)
16. Kinsler, L.E., Frey, A.R., Coppens, A.B., Sanders, J.V.: Fundamentals of Acoustics. John Wiley & Sons Inc., Chichester (2000)
17. Møller, H., Sørensen, M., Hammershøi, D., Jensen, C.B.: Head-related transfer functions of human subjects. *J. Audio Eng. Soc.* 43, 300–321 (1995)
18. Shinn-Cunningham, B.G.: Learning Reverberation: Considerations for Spatial Auditory Displays. In: Proc. of the 2000 International Conference on Auditory Display. International Community for Auditory Display (2000)
19. McGookin, D.K., Brewster, S.A.: Understanding Concurrent Earcons: Applying Auditory Scene Analysis Principles to Concurrent Earcon Recognition. *ACM transactions on Applied Perception* 1(2), 130–155 (2004)
20. Fernström, M., McNamara, C.: After Direct Manipulation–Direct Sonification. *ACM Transactions on Applied Perception* 2(4), 495–499 (2005)

Crossmodal Rhythm Perception

Maria Jokiniemi, Roope Raisamo, Jani Lylykangas, and Veikko Surakka

Tampere Unit for Computer-Human Interaction (TAUCHI)

Department of Computer Sciences
FIN-33014 University of Tampere, Finland
{forename.lastname}@cs.uta.fi

Abstract. Research on rhythm perception has mostly been focused on the auditory and visual modalities. Previous studies have shown that the auditory modality dominates rhythm perception. Rhythms can also be perceived through the tactile senses, for example, as vibrations, but only few studies exist. We investigated unimodal and crossmodal rhythm perception with auditory, tactile, and visual modalities. Pairs of rhythm patterns were presented to the subject who made a same-different judgment. We used all possible combinations of the three modalities. The results showed that the unimodal auditory condition had the highest rate (79.2%) of correct responses. The unimodal tactile condition (75.0%) and the auditory-tactile condition (74.2%) were close. The average rate remained under 61.7% when the visual modality was involved. The results confirmed that auditory and tactile modalities are suitable for presenting rhythmic information, and they are also preferred by the users.

Keywords: Crossmodal interaction, auditory interaction, tactile interaction, visual interaction, rhythm perception.

1 Introduction

Rhythms are most commonly perceived from auditory stimulation such as hearing music. Rhythms and temporal patterns can, however, also be provided through other sensory modalities, like vision and touch. Perception and recognition of rhythms has been studied extensively using auditory and visual stimuli. The results of several studies [5,6,7] show that auditory rhythms are recognized and reproduced more accurately than visual rhythms.

It has been shown that performance in rhythm comparison tasks using auditory rhythmic stimuli is frequently superior to that in tasks using visual rhythmic stimuli [4]. Glenberg et al. [6] showed through a series of experiments that the auditory superiority is not due to the alerting nature of auditory stimuli, people's greater experience with auditory rhythms, or specific response requirements in rhythm reproduction tasks. Glenberg and Jona [5] were able to diminish the auditory advantage when chunking of the beats was disturbed or long beat durations were used. Collier and Logan [4] showed that when comparing the rhythms at fast presentation rates, mixed modality rhythm pairs were as difficult as or more difficult than the pure visual rhythm pairs. At slower presentation rates, the recognition rate of the mixed modality rhythms was between the rates obtained in the unimodal visual and auditory conditions.

Rhythm perception through the tactile modality has largely been left uninvestigated. As touch is a cutaneous sense tactile information is perceived via skin. Perceiving rhythms through the sense of touch is natural. Sounds are based on similar waves as vibrotactile effects. People can, for example, feel loud music also as tactile vibrations. In user interfaces, vibrotactile rhythmic patterns can be used to present information to the user. For example, tactile icons can be used to communicate messages non-visually [1].

Tactile stimuli can provide an extension to the interaction channels between computers and visually or hearing impaired users. Vibrotactile stimuli could be used, for example, to give hearing impaired users information about events and states of a messaging device. With blind users, the auditory information channel can easily become overloaded if it is the only sensory modality used for giving feedback to the user. Thus, the use of vibrotactile stimuli could lower the pressure on the auditory information channel. The use of tactile rhythms can also support people with no sensory impairments, for example, in mobile devices, wearable computing, and learning tools [1,8].

Tactile rhythms have not been widely utilized in user interfaces, but some prototype applications have been developed. Tactons, or tactile icons, can be used to communicate complex concepts in desktop computers, in mobile and in wearable devices, and applications for visually impaired users [1]. Vibrotactile rhythms have been shown to be an effective parameter in Tactons. Brown et al. [3] evaluated a set of Tactons using three values of roughness and three different rhythms. They found an overall recognition rate of 71%, and recognition rate of 93% for rhythm. Vibrotactile effects have been successfully used to present progress information in desktop human-computer interfaces [2]. Vibrotactile rhythms have also been utilized in learning tools. The T-RHYTHM system, a rhythm instruction tool for school children, provides a child with rhythm patterns through the tactile senses. The results showed that the participants performed better when the rhythm example was given with the T-RHYTHM system than after hearing the melody played through a speaker. The system supports individual learners in playing instruments or singing, in solo or ensemble situations [8]. Kosonen and Raisamo [7] studied auditory, visual and tactile rhythm recognition in rhythm reproduction tasks. The results showed that the auditory modality dominated the tactile and visual modalities. Performance with the tactile modality was better than with the visual modality.

Crossmodal information presentation refers to a special kind of multimodal interaction where information is presented through a different sensory modality than normally expected. Crossmodal perception refers to human perception through this exceptional modality. This kind of presentation of information is useful in situations where one of the senses is temporally unavailable, such as in a noisy environment or when a person is moving and cannot look at the device. Crossmodal presentation is sometimes the only way to present information, such as pictures, for disabled people. For example, visually impaired people can investigate pictures through sonification and haptic presentation.

Our goal was to acquire basic knowledge about unimodal and crossmodal comparisons of visual, auditory, and tactile rhythms. We used pairs of the same and different rhythm patterns that were presented to the user. In rhythm presentation, we used all possible combinations of auditory, visual, and tactile modalities, which resulted in

three same-modality conditions and six crossmodality conditions. We also varied the rhythm length within each condition.

The main research question was the following: how accurately are same-different judgments of rhythms made when the rhythms are presented with different combinations of the visual, auditory, and tactile modalities? In addition, we collected users' opinions about the rhythms presented through different modalities.

2 The Experiment

2.1 Experimental Tasks

An experimental paradigm introduced by Collier & Logan [4] was used. The procedure consisted of trials where two rhythms were presented to the subject sequentially, separated by a short interstimulus interval (ISI). In a half of the trials, the rhythms were identical, and in half they were different. In each trial, the subject had to decide whether the rhythms were the same or different. The modality conditions used were the following: auditory-auditory (AA), tactile-tactile (TT), visual-visual (VV), auditory-tactile (AT), tactile-auditory (TA), auditory-visual (AV), visual-auditory (VA), tactile-visual (TV), and visual-tactile (VT).

2.2 Subjects

Twelve adults participated in the experiment (four women, eight men). Their ages ranged from 25 to 43 years (mean 30.4 yrs, median 29.5 yrs). Almost all subjects (11) had some experience with haptic devices.

2.3 Stimuli

Each rhythm consisted of five or six beats delimited by a 300-millisecond interval. We decided to use three beat lengths that were formed with the ratio 1:2:3, or 125, 250, and 375 milliseconds. For the same-pattern trials, the rhythmic pattern was presented two times in succession. For the different-pattern trials, the second pattern was altered by changing one or two elements. The rhythmic patterns used with all modality conditions are displayed in Table 1. There were 10 same-pattern trials and 10 different-pattern trials within each modality condition. Overall, there were 180 experimental trials. The patterns that had five beats ranged in duration from 2200 to 2575 ms, and the patterns that had six beats ranged from 2625 to 3125 ms. Mean trial duration was 5669 ms.

The tests were run in a usability laboratory using a standard PC computer. The auditory stimuli were played through in-ear headphones using pulses of white noise delimited by a 300-millisecond ISI. The tactile stimuli were created with a Logitech iFeel vibrotactile mouse with its vibration magnitude set in the maximum. The frequency of the vibration was 58.82 Hz. The selection of these parameters was based on user preferences in a study that investigated detection thresholds in frequency and magnitude of mouse and trackball vibration [10]. The visual stimuli were displayed

Table 1. The rhythmic patterns used in the experiment

The first pattern (the second pattern in same-pattern trials)		The second pattern in different-pattern trials	
5 beats	6 beats	5 beats	6 beats
SMSLM	MSSMLS	SLMLM	MMSMLL
LMSSL	LSLSSM	LMSSM	MSMSSM
SSMSL	SSSLML	SMMSM	SSMLMS
MMSLS	LMMLSS	SMSLS	LLSLSS
MLSMS	MMSLSS	MSSMM	MSSLMS
SMLLS	LSMSMM	SMLMS	LMMSSM

L=long beat, M=medium beat, S=short beat

on a computer screen with a square object that appeared and disappeared in the pace of the rhythm. The subject wore hearing protectors to mask the sound of the vibrotactile mouse.

2.4 Design

There were two within-subjects factors in the experiment: modality condition (AA, TT, VV, AT, TA, AV, VA, TV, and VT) and rhythm length (five and six beats). The presentation order of the modality blocks and the trials within each block were randomized separately for each subject.

2.5 Procedure

Before the experiment session, the subject was given instructions and he or she completed five practice trials. Then the subject could begin the 180 experimental trials. The researcher was present in the room during the session.

On each trial, the subject was presented with two rhythmic patterns separated by a 300-millisecond break. After both stimuli had been presented, two buttons labeled as *the same* and *different* were displayed on the screen. The subject was instructed to click one of the buttons. After each response, the next trial was immediately initiated.

Each session took about 30 minutes. The subject was allowed to take two short breaks during the session. After completing all the tasks, the subject filled in a questionnaire where he or she was asked to evaluate his or her effort, performance level, mental demand, frustration level, and preference for the modalities using a scale of 1-20 (NASA TLX [9]). After that, the subject was interviewed in more detail.

2.6 Data Analysis

Within-subject repeated measures analysis of variance (ANOVA) was used for statistical analysis. If the sphericity assumption of the data was violated, Greenhouse-Geisser corrected degrees of freedom were used to validate the F statistic. Pairwise Bonferroni corrected t -tests were used for post hoc tests.

3 Results

3.1 The Effect of Modality Conditions

The overall proportion of correct responses in distinguishing the rhythms as the same or different was 66.8 %. The mean percentages of correct responses and standard error of the means (S.E.M.) are presented in Figure 1.

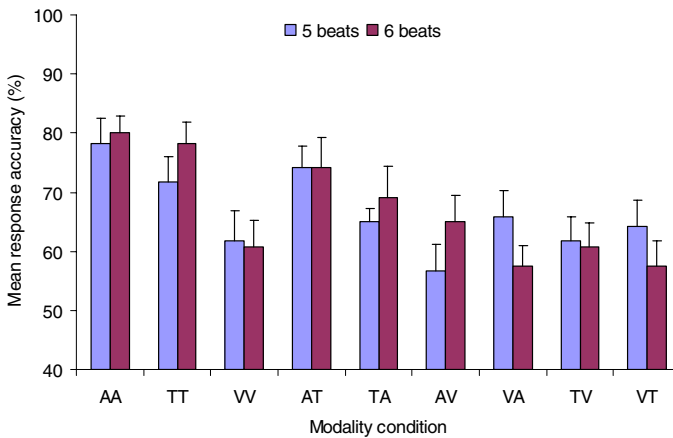


Fig. 1. Mean percentages of correct responses and S.E.M. in distinguishing sequential rhythms as the same or different

For the percentages of correct responses, a two-way 9×2 (modality condition \times rhythm length) ANOVA showed a statistically significant main effect of the modality condition ($F(8, 88) = 6.7, p < 0.001$). The main effect of the rhythm length and the interaction of the main effects were not statistically significant. Post hoc pairwise comparisons showed that subjects distinguished rhythms in AA modality condition significantly more accurately than in VA ($MD = 17.5, p < 0.05$) and TV ($MD = 17.9, p < 0.05$) modality condition. The other pairwise comparisons were not statistically significant, although the differences between AA and VV ($MD = 17.9, p = 0.083$) and between AA and AV ($MD = 18.3, p = 0.061$) modality conditions approached significance.

3.2 Subjective Ratings

For the ratings of the mental demand (see Figure 2), a one-way ANOVA showed a statistically significant effect of the modality ($F(2, 22) = 24.1, p < 0.001$). Post hoc pairwise comparisons showed that subjects evaluated the auditory modality as significantly less mentally demanding than the tactile ($MD = -4.3, p < 0.05$) and the visual ($MD = -8.9, p < 0.001$) modality. Tactile modality was evaluated as significantly less mentally demanding when compared to the visual modality ($MD = -4.6, p < 0.05$).

For the ratings of the frustration (see Figure 3), a one-way ANOVA showed a statistically significant effect of the modality ($F(2, 22) = 17.1, p < 0.001$). Post hoc pairwise comparisons showed that subjects evaluated the visual modality as significantly

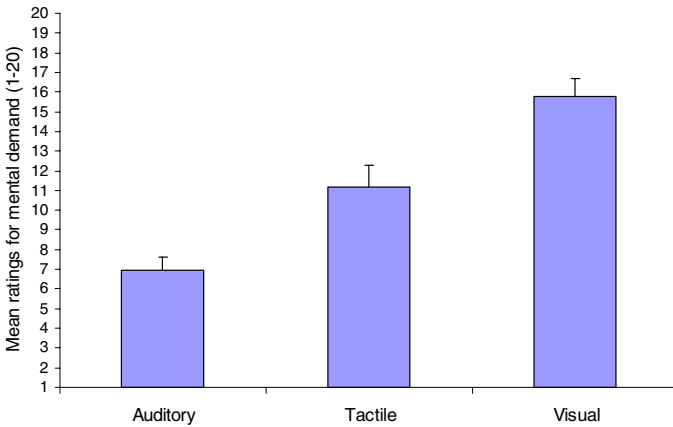


Fig. 2. Mean ratings and S.E.M. for mental demand of the modalities

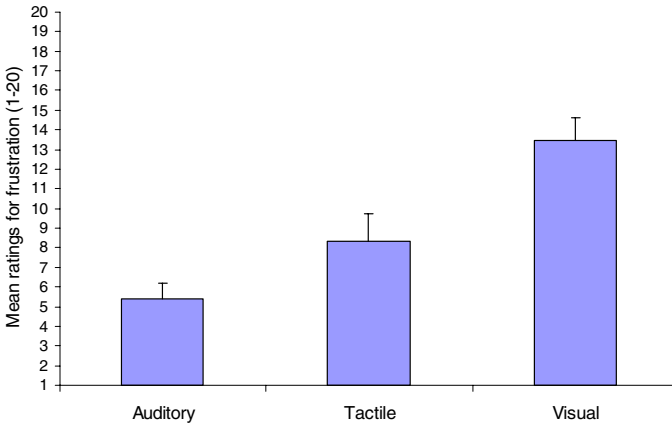


Fig. 3. Mean ratings and S.E.M. for frustration of the modalities

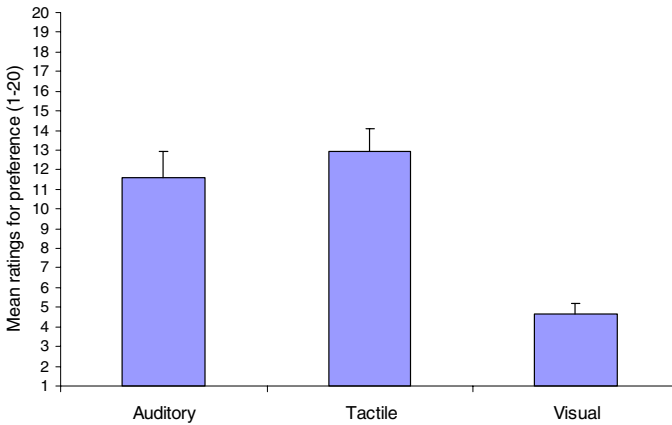


Fig. 4. Mean ratings and S.E.M. for preference of the modalities

more frustrating than the auditory ($MD = 8.0$, $p \leq 0.001$) and the tactile ($MD = 5.1$, $p < 0.01$) modality. The difference between auditory and tactile modality was not statistically significant.

For the ratings of the preference (see Figure 4), a one-way ANOVA showed a statistically significant effect of the modality ($F(2, 22) = 20.7$, $p < 0.001$). Post hoc pairwise comparisons showed that subjects evaluated the tactile modality as significantly more preferable than the visual modality ($MD = 8.3$, $p < 0.001$). Also the auditory modality was evaluated as significantly more preferable when compared to the visual modality ($MD = 7.0$, $p < 0.01$). The difference between tactile and auditory modality was not statistically significant.

Almost all subjects had negative comments on the visual rhythms. They commented that the visual modality was the most difficult one, that the visual rhythms evoked negative emotions, or that it is not a natural way to perceive rhythms. All subjects said that the VV condition was the hardest or one of the hardest conditions in the experiment. Most subjects (10 of 12) thought that the AA condition was the easiest. One subject felt that the TT condition was the easiest, and another that the AT was the easiest. Some subjects indicated that their strategy was to convert the tactile rhythms, and sometimes also the visual rhythms, into sound. None of the subjects had heard the sound related to tactile mouse vibration through the hearing protectors.

4 Discussion and Summary

The auditory superiority over the visual modality in rhythm perception has been demonstrated in several studies [2,3,4]. Our results confirm that the visual modality is the least suitable for presenting and accurately perceiving rhythmic information. Our contribution was to add the tactile modality and the related crossmodal combinations

in the experiment. Our results coincide with those of Collier and Logan [4] who showed that at fast presentation rates, auditory-visual and visual-auditory rhythm pairs were as difficult as or more difficult than unimodal visual pairs. In our experiment, all the crossmodal rhythm pairs involving the visual modality were as hard to recognize as the unimodal visual pairs. The tactile modality performed in rates closely to the auditory modality in unimodal TT and crossmodal AT conditions. Crossmodal transformation from audio to tactile was easier than from tactile to audio.

The results of subjective opinion questionnaires showed that the users clearly experienced differences in mental demand and frustration among the modalities and reported preferences among different modalities. For mental demand and frustration, the auditory modality was experienced as the least demanding modality. The visual modality was ranked as the most demanding modality, and the tactile modality was in between them. The subjects preferred the tactile modality over the auditory, but the difference wasn't statistically significant, so they can be considered to be equally preferred. These results are encouraging for the use of the tactile sense in rhythmic interaction. The results also clearly show that the users didn't like the visual rhythms.

In the future, research is needed on the information processing limits in unimodal and crossmodal rhythm perception. Comparisons need to be made using different tempos and rhythm lengths with all the three modalities and varying other parameters in rhythm presentation (e.g. the type of sound or frequency of vibration). These results can be applied for users with special needs such as for visually impaired people.

Acknowledgments. This study was carried out in the project MICOLE (IST-2003-511592 STP), funded by the European Commission. This work was also partially supported by The Finnish Agency for Technology and Innovation (Tekes), decision 40219/06. We thank all the participants of the experiment, as well as colleagues who contributed in the design and implementation of this study. Jouni Salo implemented the tactile rhythm software.

References

1. Brewster, S., Brown, L.M.: Tactons: Structured Tactile Messages for Non-Visual Information Display. In: Proc. 5th conference on Australasian user interface, vol. 28, pp. 15–23. ACM Press, New York (2004)
2. Brewster, S.A., King, A.: The Design and Evaluation of a Vibrotactile Progress Bar. In: Proc. worldHAPTICS 2005, pp. 499–500. IEEE Press, Los Alamitos (2005)
3. Brown, L.M., Brewster, S.A., Purchase, H.C.: A First Investigation into the Effectiveness of Tactons. In: Proc. worldHAPTICS 2005, pp. 167–176. IEEE Press, Los Alamitos (2005)
4. Collier, G.L., Logan, G.: Modality differences in short-term memory for rhythms. *Memory & Cognition* 28, 529–538 (2000)
5. Glenberg, A.M., Jona, M.: Temporal coding in rhythm tasks revealed by modality effects. *Memory & Cognition* 19, 514–522 (1991)
6. Glenberg, A.M., Mann, S., Altman, L., Forman, T., Procise, S.: Modality effects in the coding and reproduction of rhythms. *Memory & Cognition* 17, 373–383 (1989)
7. Kosonen, K., Raisamo, R.: Rhythm perception through different modalities. In: Proc. EuroHaptics 2006, pp. 365–370 (2006)

8. Miura, S., Sugimoto, M.: T-RHYTHM: A System for Supporting Rhythm Learning by Using Tactile Devices. In: Proceedings of IEEE International Workshop on Wireless and Mobile Technologies in Education, pp. 264–268. IEEE Press, Los Alamitos (2005)
9. NASA TLX, NASA Task Load Index,
<http://humansystems.arc.nasa.gov/groups/TLX/>
10. Raisamo, J., Raisamo, R., Kosonen, K.: Distinguishing Vibrotactile Effects with Tactile Mouse and Trackball. In: Proc. HCI 2005, pp. 337–348. Springer, Heidelberg (2005)

The Effect of Auditory Cues on the Audiotactile Roughness Perception: Modulation Frequency and Sound Pressure Level

M. Ercan Altinsoy

Chair of Communication Acoustics, TU Dresden, Helmholtzstr. 10,
01069 Dresden, Germany
ercan.altinsoy@ias.et.tu-dresden.de

Abstract. Scraping a surface with the finger tip is a multimodal event. We obtain information about the texture, i.e. roughness of the surface, at least through three different sensory channels, i.e. auditory, tactile and visual. People are highly skilled in using touch-produced sounds to identify texture properties. Sound pressure level, modulation frequency and pitch of the touch-induced scraping sounds are the important psychoacoustical determinants of the texture roughness perception. In this study, psychophysical experiments were conducted to investigate what are the relative contributions of the auditory and tactile sensory modalities to the multimodal (audiotactile) roughness percept?, what are the effects of the perceptual discrepancy between the modalities on the multimodal roughness judgment and how different modulation frequency and loudness conditions affect the subjects' roughness perception.

Keywords: Multimodal interaction, roughness, texture perception, auditory, haptic.

1 Introduction

Texture perception is an important exploration mechanism of humans to identify objects and their properties. For example, in our daily life texture information is useful to evaluate the quality of clothes, in the field of medicine, doctors use it to investigate the abnormalities of tissue in a patient.

Roughness of the surfaces is the most important physical and perceptual determinant of texture perception. Therefore most studies related to human response to textures were concentrated upon the investigation of roughness perception. The physical roughness of any surface can be defined as the height of the surface along a line across the surface. The most common and general measures of roughness are the average roughness (R_a), which is the area between the texture profile and its mean line, or the integral of the absolute value of the roughness profile height over the evaluation length, and root mean square roughness (R_q). Aside from the average roughness or the root-mean-square roughness, some other measures are used to define roughness of the particular surfaces. For example the grit numbers of the sandpaper is a measure of their physical roughness. It is a reference to the number of abrasive particles per inch of sandpaper. The lower the grit-number, the rougher the sandpaper and visa versa.

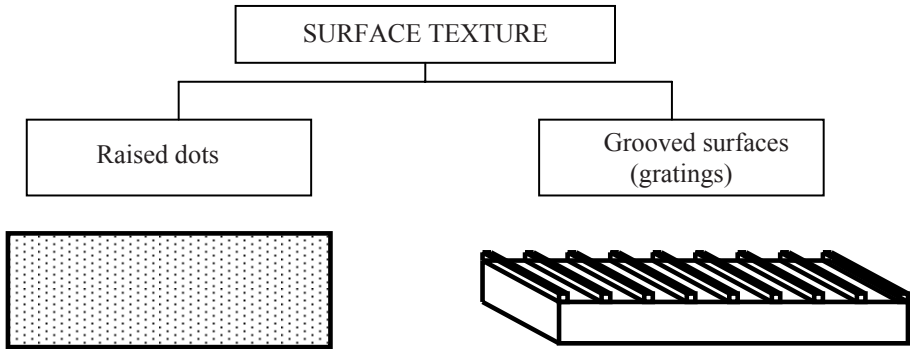


Fig. 1. Categorization of the textures for the psychophysical studies

Realistic texture profiles are mostly non-linear and randomly characterized, therefore, to eliminate the difficulties (analysis) and to control the conditions, in most psychophysical studies (i.e. linear) surfaces were used. Regarding the psychophysical studies on the roughness perception, textures can be categorized and simplified into two different stimulus-categories: raised dots, e.g. abrasive surfaces such as sandpaper etc. and grooved surfaces, e.g. grammophone plaque (Figure 1).

In this study, the grooved surfaces were selected as stimuli. During the fingertip scraping across the surface of the grooved wood, the ridges experience the force which is applied by the fingertip and their movement is transmitted to the block. The vibrations of the wooden block and the ridges are the predominant sources of the noise. This noise is modulated by the contact of the fingertip with the ridges. The frequency of the modulation is proportional to the ridge number and the scraping velocity and can be described as follows;

$$fr1 = v r / L \quad (1)$$

where r is the total ridge number, L is the length of the block and v is the scraping velocity.

1.1 Multimodal Roughness Perception

People are capable of evaluating the roughness of surfaces moved across their fingertips. The investigation of the relationship between physical-roughness-descriptors and roughness perception is an interesting research topic for virtual-environment designers who want to mimic different textures in their environment.

In a study by Lederman [4], congruent auditory and tactile texture information was presented to the subjects and they were asked for the roughness of the surfaces. She found that, if tactile and auditory sources of information are available, subjects tend to use tactile cues to judge surface roughness. This result indicates that tactile texture cues completely dominate the auditory cues in determining texture perception. Her explanation for this result is that in daily life sound cues, which are generated by touching the texture of a surface, are masked by background noises due to their low

level. Therefore, our attention is directed to the tactile modality. This argument was somehow confirmed in another study by Lederman et al. [7]. They experimentally assessed the relative contributions of tactile and auditory information to bimodal judgments of surface roughness using a rigid probe. The sounds generated due to contact between a rigid probe and a rigid surface are louder than those generated by bare finger. Their results indicated that when a subject explores the surfaces by a rigid probe, she/he uses both tactile and auditory information to make their estimates.

Under certain conditions, auditory cues which are generated by bare finger can also influence tactile roughness judgments [3]. Jousmäki and Hari have named this effect as “parchment-skin illusion”. In their experiment, subjects have rubbed their hands together and listened simultaneously to modified sounds as generated by rubbing. After the stimuli presentation, they were asked to rate roughness and moistness of the palmar skin of their hands. The results showed that when overall sound pressure level increased (20 dB or 40 dB), or when the frequency components within the frequency range of 2-20 kHz were amplified, subjects have felt smoother and dryer. If the sound pressure level decreased, they felt rougher and moister. Later on Guest et al. [2] has demonstrated that the same effect is valid for the sandpaper stimulus also.

1.2 The Objectives of the Present Study

The studies of Jousmäki and Haki [3] and Guest et al. [2] indicate that under certain conditions, auditory and tactile information can interact by determining the roughness of the textures. Increasing loudness can result in a decrease in perceived roughness. From the view of a virtual environment designer, the following question arises: If loudness can play such a role on the multimodal roughness perception, what can be the influence of auditory modulation frequency (fundamental frequency related ridge number) on the multimodal roughness perception.

Perceiving the texture of a surface by touching it (scraping with the fingertips) is a multimodal task in which information from auditory, tactile and visual sensory channels are available. What are the relative contributions of the various systems (tactile, auditory, visual) on the multimodal percept, how does incongruent sensory information interact and how can the combination of multimodal output of information be designed better?

In order to achieve these aims, experiments with unimodal and multimodal stimulus presentations were conducted and, especially, the effects of the perceptual discrepancy between the auditory and the tactile sensory modalities on the multi-sensory roughness judgment were investigated.

2 Experiments

The aim of the first experiment was to investigate the relative contributions of the auditory and tactile information on the bimodal judgments if sound pressure levels are 10 dB higher than physically accurate. Taking into account the statement of Lederman[4] which is *“In daily life sound cues, which are generated by touching the texture of a surface, are masked by background noises due to their low level. Therefore tactile texture cues completely dominate the auditory cues in determining texture*

perception”, the sound pressure levels of the auditory stimuli were amplified 10 dB as compared to the physically accurate value. The second experiment was conducted to investigate the influence of the modulation frequency information on the tactile roughness judgments.

2.1 Auditotactile Roughness Perception and Relative Contributions of the Auditory and Tactile Systems

Subjects

Ten subjects, four men and six women, aged between 22 and 29 years, participated in this experiment. The subjects were undergraduate students and paid on an hourly basis. All subjects were right handed, with no known heart and hand disorders and they used their right hand for the experiment. All subjects had self-reported normal hearing.

Experimental Set-Up

By representing the tactile texture information, electro-tactile stimulation technique was selected according to its advantages. Its operation doesn't cause any noise, this makes it especially suitable for auditory-tactile virtual environment applications. Self-adhesive electrodes were used to excite the user's fingertip. Current magnitude (mA) and pulse frequency (Hz) of the electrotactile stimulus are the parameters which allow to represent the texture profiles for different roughnesses. (Detailed information see [1]).

The auditory stimulus was presented from a PC. It was amplified and delivered diotically through Sennheiser HDA 200 closed-face dynamic headphones. The experiments were conducted in a sound-attenuated room. In order to control auditory attributes of the scraping sounds, they were synthesized in a computer environment.

Stimuli and Procedure

The stimuli were tactile information and sounds such as those generated by touching rectangular wood pieces, 14 x 4 x 1.5 cm, each with a set of linear grooves (0.25, 0.5, 0.75, 1.00, 1.50 mm) and constant 1.00 mm ridge width.

The virtual textures were presented and roughness was estimated using an absolute estimation method [9]. The subjects task was to report the degree of perceived roughness using numbers. For the first stimulus, they were asked to assign any positive, non-zero number (decimal, fraction or whole-number) that they think to be appropriate. For the next stimulus, they were required try to give an appropriate number in relation to the previous stimulus (rational). In other words if the texture feels 2 times as rough as the previous stimulus, they should assign a number which is two times the number which they had assigned to the previous stimulus, e.g. if they assigned the number 5 for the previous stimulus, they should now assign 10. The subjects were asked not to worry about being consistent.

In the training phase, which took about 15 minutes, firstly all participants were presented with different stimulus combinations from across the full stimulus range, and then they were familiarized with the magnitude-estimation procedure using six different

stimulus combinations. To prevent participants devising a fixed response range, they were informed that they might experience rougher or smoother stimuli in the actual experiment than in the training (as in [6]). In the actual experiment, each stimulus was presented in random order and four times.

Results

Roughness judgments for the conditions: auditory only, tactile only and auditory and tactile together are shown in Figure 2 as a function of the log groove width.

The data points represent log magnitude estimates and are based on 100 responses. To eliminate the influence of the chosen numerical scale (which could be freely selected by the subjects), the resulting mean magnitude estimates (each computed from participant's ten magnitude estimates) were subsequently normalized by dividing each score by the individual participant mean, then multiplying it by ten. Responses are normalized to the value 10 for 0.25 mm groove width.

Dependent t-tests of the means show that all three conditions differed significantly (auditory only – tactile only: $t(9) = -5.64$, $p < 0.05$; tactile only – audiotactile: $t(9) = 6.74$, $p < 0.05$; auditory only – audiotactile: $t(9) = -5.85$, $p < 0.05$).

To calculate a measure of the relative contributions of tactile and audition conditions to the bimodal (auditory & tactile) estimates, a technique, which is proposed by [7] for the multi-sensory perception of the surface roughness, was used. This statistic, which indicates the percent weighting of the tactile information in the bimodal judgments, was calculated related to the distances between three different conditions:

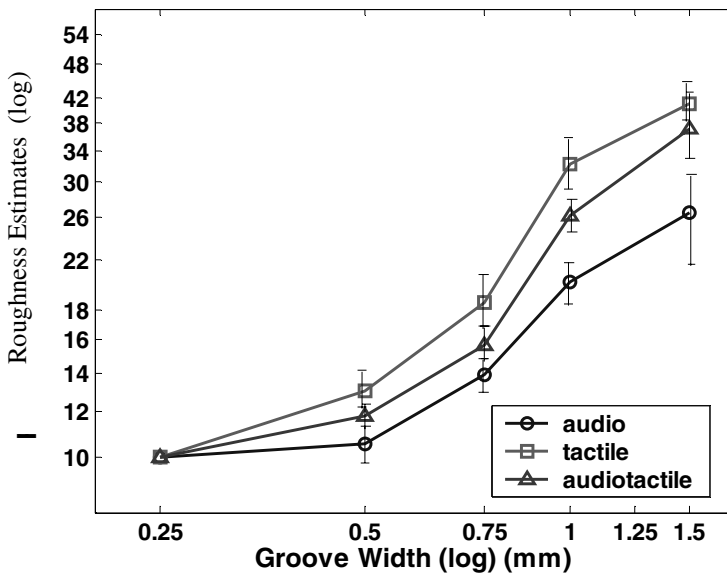


Fig. 2. Perceived roughness as a function of groove width and sensory mode of the judgment (audio, tactile, audiotactile). The data are averaged across subjects.

$$\% T_{\text{dominance}} = \left[\frac{(\text{Mean}_{\text{Tactile+Audio}} - \text{Mean}_{\text{Audio only}})}{(\text{Mean}_{\text{Tactile only}} - \text{Mean}_{\text{Audio only}})} \right] \quad (2)$$

The relative weighting of the tactile only condition is approximately 60% and the relative weighting of auditory only condition is approximately 40%.

Discussion

In all three conditions, subjects could judge the roughness of wooden plates for varying groove width. Perceived roughness increases with increasing groove width. In all three conditions the roughness estimates differed from each other.

The slope of the auditory only condition shows slower acceleration as seen in the other conditions. This result is in line with the results of Lederman [4]. In the tactile only condition, the roughness estimates are higher than in the auditory-only and auditory-and-tactile-together conditions.

The curve of the auditory-tactile roughness judgments and the results of the relative weightings show that the subjects take into account both tactile and auditory information. These results do not agree with the results of the Lederman [4], who found that touch based auditory cues do not play any role on the bimodal judgments. Recall that she has argued that low-level sound cues are frequently masked by the general background noise in many everyday situations. One of the reasons for the difference between the results of the present study and the results of Lederman [4] could be that in the present study all sound-pressure-levels are 10 dB above the physically accurate value and this amplification results in an increase of the contribution of the auditory information on the bimodal judgments. The results of Lederman et al. [7] on the assessment of bimodal roughness judgments using a rigid probe confirm this argument. With rigid contact between surface and end effector, the amplitude of the accompanying sounds is usually considerably greater and their results show that the subjects used not only tactile information, but also auditory information on the bimodal judgments. These results also confirm the results of the Jousmäki and Hari [3], and Guest et al. [2] that under certain conditions auditory cues which are generated by the bare finger can also influence tactile roughness judgments.

2.2 Influence of Modulation Frequency on Roughness Perception

The aim of this experiment was to investigate the role of the modulation-frequency information on the tactile-roughness judgments. Subjects, set-up and procedure were the same as in the first experiment. In this experiment, some congruent (modulation frequency and tactile frequency) and incongruent stimuli (Table 1) pairs were presented.

Similarly to other experiments, the absolute-magnitude-estimation method was used in this experiment. The subject's task was to report how rough they felt by assigning numbers regarding the roughness of the tactile stimulus. They were specifically instructed to ignore the touch sounds they heard, and to base their judgments only on tactile information.

Table 1. Stimuli list of the experiment

Stimuli Number	Auditory Stimulus	Tactile Stimulus
1	0.25 mm groove width (mod. freq. 112 Hz)	0.5 mm groove width
2	0.5 mm groove width (mod. freq. 94 Hz)	0.5 mm groove width
3	1 mm groove width (mod. freq. 70 Hz)	0.5 mm groove width
4	0.25 mm groove width (mod. freq. 112 Hz)	0.75 mm groove width
5	0.75 mm groove width (mod. freq. 80 Hz)	0.75 mm groove width
6	1.5 mm groove width (mod. freq. 56 Hz)	0.75 mm groove width

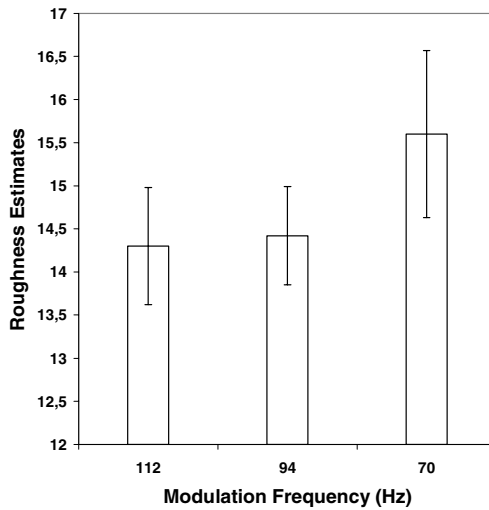


Fig. 3. Perceived roughness of the stimuli 1, 2, 3 (see Table 1). The data are averaged across subjects.

Results

The roughness estimates (and SE's) for the stimulus numbers 1, 2, and 3, as a function of the auditory modulation frequency (groove width) are shown in Figure 3. Figure 4 shows the data for the stimulus numbers 4, 5, and 6, as a function of auditory modulation frequency. The data points represent geometrical means of the 100 responses.

Discussion

The results show that in incongruent stimuli presentations, the auditory modulation frequency can alter the tactile information. Decreasing modulation frequency results in an increase in perceived tactile roughness, even though tactile information is

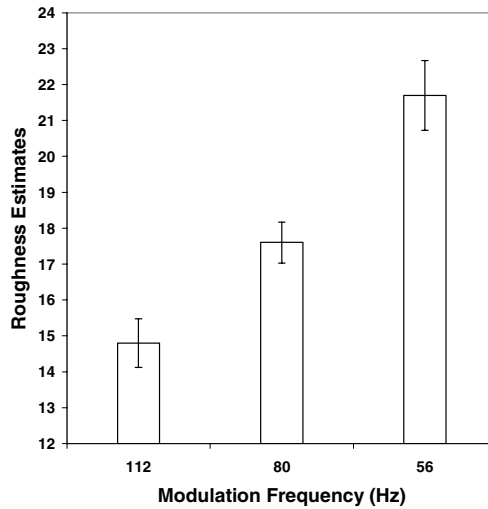


Fig. 4. Perceived roughness of the stimuli 4, 5, 6 (see Table 1). The data are averaged across subjects.

smoother than the auditory information. This effect is also observable for the increasing-modulation-frequency condition. Here the increasing modulation frequency results in a decrease of the roughness estimate, even though tactile channel indicates that it is rougher. The effect can be seen very clearly for the stimulus numbers 4, 5 and 6, but less clear for the stimulus numbers 1, 2, and 3. One reason may be that the small difference between 0.5 mm and 0.25 mm groove width is not enough to alter the tactile information.

3 General Discussion and Conclusions

Texture perception of a surface by touch is a complex, multimodal process, and it is difficult to simulate realistic multimodal textures using virtual-reality displays. Generation of multimodal textures for virtual reality applications also requires knowledge about the nature of the realistic and virtual textures. In this study, the design of multimodal textures for virtual-reality applications was discussed by investigating unimodal and multimodal (auditory-tactile) roughness-perception issues.

The results of the multimodal roughness judgment experiment show that subjects take both tactile information and auditory information into account in their bimodal judgments. The auditory modality (40 % weighting in the bimodal judgments) is nearly as informative regarding the roughness of the textures as the tactile modality is (60 % weighting in the bimodal judgments). Auditory attributes, e.g. modulation frequency and loudness, can alter significantly the tactile-roughness-perception in bimodal stimulus presentation conditions, if they are incongruent with the tactile

information. Decreasing modulation frequency results in an increase in the perceived tactile roughness as in some conditions (small increment, such as 6 or 9 dB) increasing loudness results an increase in the perceived tactile roughness, while, on the contrary, in other conditions (great increment, such as 20 or 40 dB) increasing loudness results in a decrease in the perceived tactile roughness. Interaction on the tactile roughness perception related to auditory and tactile attributes is complex, and the effects depend on the conditions. Therefore, designers should be aware of this complexity if they want to use benefits of the bimodal stimuli presentation. The results of the study indicate that the auditory information can be useful in overcoming the limitations of the haptic texture presentation techniques and provide further realism.

When the results of the multimodal roughness judgment experiments are interpreted from the intersensory-organization perspective, the modality superiority hypothesis, as is suggested by Welch et al. [8] and Lederman, and Abbot [5] is in line with the current findings. The measures of the modality superiority hypothesis are accuracy, sensitivity, discrimination, precision, and other aspects of performance. The results of the current study show that if the sound pressure level of the scraping sounds is 10 dB above the physically accurate value, auditory and touch performances related to roughness of the textures are very similar. Auditory judgments are nearly as precise and discriminative as the touch judgments, and information can be obtained easily and quickly by both modalities. Therefore both information were used by the subjects and somehow they superimpose both information in their bimodal judgments. This argumentation can be confirmed by another intersensory-organization hypothesis, namely, ecological validity, which is suggested by Lederman [4]. Lederman argued that one reason that tactile sense may bias auditory sense in a texture-related task, (in situations where audition is less discriminating than touch) is that tactual cues to texture are more ecologically valid than auditory cues. In the current experiment the auditory sense was nearly as influential as touch when judging the roughness of surfaces, therefore the task (roughness judgment) may be considered ecologically valid for both modalities.

References

1. Altinsoy, E.: Auditory-Tactile Interaction in Virtual Environments. Shaker Verlag, Germany (2006)
2. Guest, S., Catmur, C., Llyod, D., Spence, C.: Audiotactile Interactions in Roughness Perception. *Exp. Brain Res.* 146, 161–171 (2002)
3. Jousmäki, V., Hari, R.: Parchment-skin Illusion: Sound-biased Touch. *Touch. Current Biology* 8, 190 (1998)
4. Lederman, S.J.: Auditory Texture Perception. *Perception* 8, 93–103 (1979)
5. Lederman, S.J., Abbott, S.G.: Texture Perception: Studies of Intersensory Organization Using a Discrepancy Paradigm and Visual Versus Tactual Psychophysics. *J. Exp. Psychol: Human Perception and Performance* 7, 902–915 (1981)
6. Lederman, S.J., Klatzky, R.L., Hamilton, C.L., Ramsay, G.I.: Perceiving Roughness via a Rigid Probe: Psychophysical Effects of Exploration Speed and Mode of Touch. *Haptics-e (Electronic Journal of Haptics Research)* 1, 1–20 (1999)

7. Lederman, S.J., Klatzky, R.L., Hamilton, C., Morgan, T.: Integrating Multimodal Information About Surface Texture via a Probe: Relative Contributions of Haptic and Touch Produced Sound Sources. In: 10th Annual meeting of Haptic Interfaces for Teleoperator and Virtual Environment Systems. Satellite meeting of the Annual IEEE VR 2002 meeting, pp. 97–104 (2002)
8. Welch, R., Widawski, M., Harrington, J., Warren, D.: An Examination of the Relationship Between Visual Capture and Prism Adaptation. *Percept. Psychophys.* 25, 126–132 (1979)
9. Zwislocki, J., Goodman, D.: Absolute Scaling of Sensory Magnitudes: A Validation. *Percept. Psychophys* 28, 28–38 (1980)

Author Index

- Altinsoy, M. Ercan 120
Avizzano, Carlo Alberto 30
- Bergamasco, Massimo 30
Boll, Susanne 1
- Carriço, Luís 60
Chang, Angela 70
- de Sá, Marco 60
Devallez, Delphine 102
- Fontana, Federico 102
- Goncharenko, Igor 40
- Hall, Richard 50
Harris, Peter 50
Henze, Niels 1
Hermann, Thomas 91
Heuten, Wilko 1
- Ioannou, Ioanna 50
- Jokiniemi, Maria 111
- Kazmierczak, Edmund 50
Kõiva, Risto 91
Kryssanov, Victor V. 40
Kumokawa, Shizuka 40
- Leonardi, Rosario 30
Lylykangas, Jani 111
- MacLean, Karon 21
Maiorca, Mauro 50
- O'Leary, Stephen 50
O'Sullivan, Conor 70
Ogawa, Hitoshi 40
- Pasto, Virpi 11
Pedrosa, Ricardo 21
Pielot, Martin 1
Pirhonen, Antti 81
Portillo-Rodriguez, Otniel 30
- Raisamo, Roope 11, 111
Rathod, Hemang 50
Reis, Tiago 60
Rocchetto, Davide 102
Ruffaldi, Emanuele 30
- Sallnäs, Eva-Lotta 11
Sandoval-Gonzalez, Oscar O. 30
Surakka, Veikko 111
- Tanhua-Piiroinen, Erika 11
Tuuri, Kai 81